

1-1-2005

# Desire-satisfaction theories of welfare.

Christopher C. Heathwood  
*University of Massachusetts Amherst*

Follow this and additional works at: [https://scholarworks.umass.edu/dissertations\\_1](https://scholarworks.umass.edu/dissertations_1)

---

## Recommended Citation

Heathwood, Christopher C., "Desire-satisfaction theories of welfare." (2005). *Doctoral Dissertations 1896 - February 2014*. 2395.  
[https://scholarworks.umass.edu/dissertations\\_1/2395](https://scholarworks.umass.edu/dissertations_1/2395)

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).



# **DESIRE-SATISFACTION THEORIES OF WELFARE**

A Dissertation Presented

by

CHRISTOPHER C. HEATHWOOD

Submitted to the Graduate School of the  
University of Massachusetts Amherst in partial fulfillment  
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2005

Philosophy

© Copyright by Christopher C. Heathwood 2005

All Rights Reserved



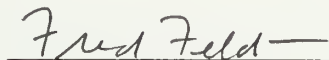
# DESIRE-SATISFACTION THEORIES OF WELFARE

A Dissertation Presented

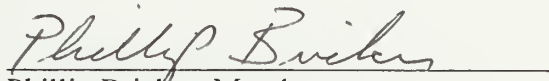
by

CHRISTOPHER C. HEATHWOOD

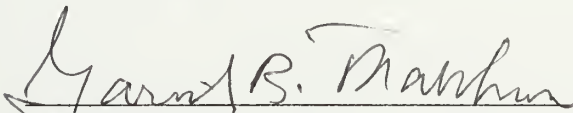
Approved as to content and style by:



Fred Feldman, Chair



Phillip Bricker, Member



Gareth Matthews, Member



Angelika Kratzer, Member



Phillip Bricker, Department Head  
Philosophy

## **DEDICATION**

To Mom and Dad, and Nicki and Henry,  
the people who are responsible for most all the good in my life.

## ACKNOWLEDGMENTS

Were it not for the guidance, support, inspiration, and encouragement of many generous and talented people, I would not have come to study philosophy in graduate school and would not have completed this dissertation. I am extremely grateful to each of these people for his or her contribution to my coming to be a professional philosopher.

I am grateful to my teachers at the University of California, San Diego, where I was first exposed to philosophy and where I studied it as an undergraduate. I am especially grateful to Patricia Kitcher, who met with me weekly for two trimesters and guided me as I tried to write an honors thesis on Hume; to Paul Churchland, in whose class I fell in love with contemporary analytic philosophy; and to Michael Mendelson, whose wonderful teaching in my first philosophy courses prompted me to become a philosophy major. I also want to thank the graduate student “reader” in the course I took with Professor Churchland. The very kind praise he gave me on my term paper helped me to think it might not be so crazy for me to apply to graduate school in philosophy.

I owe an unusual debt of gratitude to my speech therapist, Dr. Edmund Thile. Because I stutter, I was convinced that I could never land a job as a university professor. Without Dr. Thile’s assistance, I am fairly confident I would never have applied to graduate school in philosophy and would never have had the courage to start teaching as a graduate student instructor.

I am grateful to my teachers at the University of Massachusetts, Amherst; they have shaped me into the philosopher I am today. The teachers most responsible for whatever philosophical skills I may have today are Fred Feldman and Phil Bricker. I took or sat in on just about every course taught by Fred or Phil.

I also want to thank Phil Bricker, Angelika Kratzer, and Gary Matthews for serving as Members of my dissertation committee, and for assisting me along the way with the project.

Early versions of Chapters Two, Three, and Five were presented at meetings of Fred Feldman's "Dissertators Club." I thank club members Ty Barnes, Dan Doviak, Kristen Hine, Justin Klockseim, Julie Petty, Jason Raibley, Michael Rubin, Jean-Paul Vessel, and especially Fred Feldman for their helpful criticism and advice during those meetings, and also for providing something most all dissertators need: deadlines.

Material from this dissertation has been presented at the University of Massachusetts, Amherst, the 2003 Pacific Division Meeting of the American Philosophical Association, the 2004 Central Division Meeting of the American Philosophical Association, Stockholm University, McGill University, Arizona State University, Michigan State University, Western Washington University, and the University of Colorado at Boulder. Many people provided outstanding feedback on each of those occasions. I thank all of these people, and especially my APA commentators, Robert Epperson and Mark van Roojen.

Chapter Two, "The Problem of Defective Desires," will appear this December (2005) in the *Australasian Journal of Philosophy*. I thank the *Australasian Journal of Philosophy* and Routledge for permission to include it here. Chapter Five, "Desire



Satisfactionism and Hedonism,” will appear in *Philosophical Studies*. I reprint it here with kind permission from Kluwer Academic Publishers.

I have benefited tremendously from countless conversations I have had about my views and about philosophical matters generally with my friends in graduate school. In this category, Kris McDaniel has been singularly brilliant. I am also extremely grateful to Jake Bridge, Jeremy Cushing, Dan Doviak, Jack Hanson, Kristen Hine, Jens Johansson, Justin Klockseim, Marcy Lascano, Uri Leibowitz, Steve Masterman, Kirk Michaelian, Jason Raibley, Creighton Rosental, Michael Rubin, Stephan Torre, Brandt Van der Gaast, and Jean-Paul Vessel.

I would not be about to earn a doctorate in philosophy if it were not for the support, in so many ways, of my wife, Nicole. Nicki encouraged me to apply to graduate school, helped tremendously in the application process, has had a “real job” to support us as I indulge my academic pursuits, and has helped me to grow up enough to do all that needs to be done to succeed in graduate school, including especially this dissertation. I simply could not have done it without her.

In a similar vein, I want to thank my parents, Desmond and Ann Heathwood. Although both of them pursued more practical careers, not once did they try to dissuade me from trying to make a career out of philosophy. Quite the contrary: they have been nothing short of 100% supportive of my aspirations to become a philosopher. The fact that what I do makes them proud has helped give me the courage to follow my dreams.

Finally, but foremost, I thank my advisor, Fred Feldman. The idea for this dissertation arose out of Fred’s seminar on “The Good Life” in the Spring of 2001. Fred read and provided invaluable feedback on many versions of every chapter of this

dissertation, and he always did it astonishingly quickly, despite being continuously busy with his own projects. Fred was always ready to help me improve the project, suggesting ways to develop and defend it, despite the fact that he is not sympathetic with it. That is a sign of an ideal advisor.

It is impossible to convey in this short space the tremendous impact Fred has had on my philosophical development. In my first semester in graduate school, I decided to use Fred as my philosophical role model, and I continue to do so to this day. Fred taught me how to write philosophy, how to read philosophy, how to formulate a philosophical thesis, and how to refute one. Just as I cannot convey Fred's impact on me as a philosopher, I likewise cannot convey how grateful I am to Fred for it. Nothing guarantees that graduate students will find an advisor who is an acceptable match for them, let alone an excellent one. But Fred's amazing dedication to his students, his unbelievable generosity with both his time and his mind, and his unsurpassed skills as a teacher and advisor just about guaranteed that I would find a perfect match. Thanks, Fred, for making my experience in graduate school exceed my wildest expectations.

## ABSTRACT

DESIRE-SATISFACTION THEORIES OF WELFARE

SEPTEMBER 2005

CHRISTOPHER C. HEATHWOOD

B.A., UNIVERSITY OF CALIFORNIA, SAN DIEGO

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Fred Feldman

Theories of welfare (or well-being or “the good life”) answer the ancient question, What makes a person’s life go well? Prominent among these are desire-satisfaction or preferentist theories, according to which welfare has to do ultimately with desire. This dissertation aims (i) to criticize some recent popular arguments against standard desire-satisfaction theories of welfare, (ii) to develop and defend a novel version of the desire-satisfaction theory capable of answering the better objections, (iii) to defend the thesis that pleasure is reducible to desire, and (iv) to demonstrate an interesting link between preferentism and hedonism.

The second chapter (the first is the introduction) defends a simple “actualist” desire-satisfaction theory against the contention that such a theory cannot accommodate the fact that we can desire things that are bad for us. All the allegedly defective desires, I attempt to show, are either not genuinely defective or else can be accounted for by the theory.

The third chapter criticizes the popular line that standard desire-based theories of welfare are incompatible with the conceptual possibility of self-sacrifice. I show that

even the simplest imaginable, completely unrestricted desire theory is compatible with self-sacrifice, so long as it is formulated properly.

The fourth chapter presents and defends a theory according to which welfare consists in the perceived satisfaction, or “subjective satisfaction,” of desire. I argue that this theory is best suited to deflect the many lines of argument threatening the preferentist program.

The fifth chapter defends the view that desire is what unifies the heterogeneous lot of experiences that all count as sensory pleasures. I develop and defend a desire theory of sensory pleasure.

The sixth chapter argues that the most plausible form of preferentism is equivalent to the most plausible form of its main rival, hedonism. This is because what the best preferentism says – that welfare consists in subjective desire satisfaction – is the same as what the best hedonism says – that welfare consists in propositional pleasure – given a reduction of pleasure to desire along the lines of that defended above.



## TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS .....	v
ABSTRACT .....	ix
Chapter	
1. INTRODUCTION .....	1
1.1 The Concept of Welfare .....	1
1.2 Theories of Welfare .....	5
1.3 Why Some Are Attracted to a Desire Approach to Welfare .....	9
1.4 Some Metaethical Remarks .....	13
2. THE PROBLEM OF DEFECTIVE DESIRES .....	15
2.1 The Problem of Defective Desires .....	15
2.2 The Actual Desire-Satisfaction Theory .....	18
2.3 The Form of The Argument .....	20
2.4 Arguments from All-Things-Considered Defective Desires .....	21
2.4.1 The Argument from Ill-Informed Desires .....	22
2.4.2 The Argument from Irrational Desires .....	25
2.4.3 The Argument from Artificially Aroused Desires .....	26
2.4.4 The Argument from Poorly Cultivated Desires .....	28
2.5 Arguments from Intrinsically Defective Desires .....	29
2.5.1 The Argument from Base Desires .....	30
2.5.2 The Argument from Pointless Desires .....	36
2.5.3 The Argument from Poorly Cultivated Desires .....	40
2.6 The Argument from Desires To Be Badly Off .....	41
2.6.1 The Paradoxes of Desire .....	42
3. DESIRE-BASED THEORIES OF WELFARE AND THE POSSIBILITY OF SELF-SACRIFICE .....	45
3.1 Preliminaries: Theories of Welfare; The Notion of Self-Sacrifice .....	47

3.2	One Kind of Desire-Based Theory of Welfare and The Argument from Self-Sacrifice . . . . .	48
3.3	Another Kind of Desire-Based Theory of Welfare . . . . .	52
3.4	Simple Desire Satisfactionism and the Possibility of Self-Sacrifice . . . . .	59
3.5	Objections and Replies . . . . .	61
3.6	Closing Remarks about Simple Desire Satisfactionism . . . . .	67
4.	SUBJECTIVE DESIRE SATISFACTIONISM . . . . .	69
4.1	Holistic Desire-Based Theories of Welfare . . . . .	70
4.1.1	Naive Preferentism and the Arguments from Ignorance, Weakness of Will, and Unimaginativeness . . . . .	71
4.1.2	Ideal Preferentism . . . . .	73
4.1.3	The Argument from Self-Sacrifice . . . . .	75
4.1.4	Egocentric Ideal Preferentism and Its Difficulties . . . . .	76
4.1.5	The Argument from Self-Loathing . . . . .	80
4.1.6	The Argument from Changing Preferences . . . . .	81
4.1.7	The Argument from Unwanted Ideal Lives . . . . .	84
4.1.8	The Argument from Unworthy Desires . . . . .	87
4.2	Atomistic Desire-Based Theories of Welfare . . . . .	89
4.2.1	Simple Desire Satisfactionism . . . . .	89
4.2.2	The Argument from Changing Desires . . . . .	93
4.2.3	An Argument from Anti-Paternalism . . . . .	95
4.2.4	The Argument from Remote Desires . . . . .	97
4.2.6	The Argument from Posthumous Harm . . . . .	100
4.3	Subjective Desire Satisfactionism . . . . .	100
4.3.1	The Experience Requirement . . . . .	100
4.3.2	Subjective Desire Satisfactionism: The Atoms of Welfare . . . . .	102
4.3.3	Subjective Desire Satisfactionism: The Value of the Atoms . . . . .	108
4.3.4	Subjective Desire Satisfactionism: The Value of a Life . . . . .	110
4.4	Subjective Desire Satisfactionism at Work . . . . .	111
4.4.1	Subjective Desire Satisfactionism and the Arguments from Ignorance, Weakness of Will, and Unimaginativeness . . . . .	111
4.4.2	Subjective Desire Satisfactionism and the Argument from Self-Sacrifice . . . . .	114
4.4.3	Subjective Desire Satisfactionism and the Argument from Self-Loathing . . . . .	114
4.4.4	Subjective Desire Satisfactionism and the Argument from Changing Desires . . . . .	115

4.4.5	Subjective Desire Satisfactionism and the Arguments from Unwanted Ideal Lives and Unwanted Ideal Desiderata . . . . .	116
4.4.6	Subjective Desire Satisfactionism and the Argument from Posthumous Harm . . . . .	117
4.4.7	Subjective Desire Satisfactionism and the Argument from Remote Desires . . . . .	118
4.4.8	Subjective Desire Satisfactionism and the Argument from Unworthy Desires . . . . .	118
4.4.9	Subjective Desire Satisfactionism and an Argument from Anti-Paternalism . . . . .	119
4.5	Concluding Remarks . . . . .	121
5.	THE REDUCTION OF SENSORY PLEASURE TO DESIRE . . . . .	123
5.1	Introduction . . . . .	123
5.1.1	Kinds of Connection Between Pleasure and Desire . . . . .	123
5.1.2	A Theory of Pleasure, and Why . . . . .	124
5.1.3	Desire . . . . .	126
5.2	The Heterogeneity Problem . . . . .	127
5.3	The Desire Approach to Pleasure and “the Oppositeness Problem” . . . .	129
5.4	Three Categories of Pleasure . . . . .	131
5.5	The Reduction of Sensory Pleasure to Desire . . . . .	133
5.6	Arguments Against the Reduction of Sensory Pleasure to Desire . . . . .	138
5.6.1	Arguments from the Concept of Desire . . . . .	138
5.6.2	An Argument from Desired Non-Pleasures: The Dizziness Case . . . . .	141
5.6.3	A Second Argument from Desired Non-Pleasures: “Interesting” Sensations . . . . .	145
5.6.4	“I want it because it’s pleasant,” or: A Euthyphro Problem . . .	148
5.6.5	An Argument from Undesired Non-Pains . . . . .	152
6.	DESIRE SATISFACTIONISM AND HEDONISM . . . . .	160
6.1	Desire Satisfactionism . . . . .	161
6.1.1	The Problem of Changing Desires . . . . .	162
6.1.2	The Objection from Remote Desires . . . . .	164
6.1.3	The Problem of Desiring Not To Be Well-Off . . . . .	165
6.1.4	The Problem of Defective Desires . . . . .	167
6.2	Subjective Desire Satisfactionism . . . . .	171
6.3	Hedonism . . . . .	176

6.4 The Motivational Theory of Pleasure .....	182
6.5 Two Theories or One? .....	186
BIBLIOGRAPHY .....	188



# CHAPTER 1

## INTRODUCTION

This dissertation is a collection of five self-contained papers (not including this introduction). The first three (Chapters 2, 3, and 4) concern desire-satisfaction theories of welfare, Chapter 5 concerns the nature of sensory pleasure, and Chapter 6 concerns the relationship between desire theories and hedonistic theories of welfare. One advantage of this format is that the reader can begin anywhere he likes (the abstract above contains very brief synopses of each chapter). A disadvantage is that the work repeats itself occasionally.

The goal of this introduction is to address some preliminary and foundational issues that are not discussed in the chapters to come. I make an effort to identify the concept of welfare while explaining its importance. I lay out what I take a theory of welfare to be. I outline some of the reasons many thinkers are attracted to a desire approach to welfare. I conclude the introduction with some brief metaethical confessions.

### 1.1 The Concept of Welfare

Among the greatest and oldest of questions in the history of western philosophy is, What makes a person's life go well for him? In other words, What things are good in themselves for a person? We ask the same question when we wonder in what human welfare consists.

It is an understatement to say that the concept of welfare plays a prominent role in moral philosophy. It plays center stage in utilitarian theories of right and wrong, standard versions of which require us to maximize the amount of welfare in the world. Theories of conduct of a more pluralistic bent must also make use of the notion of welfare, for, as I see it, any plausible normative theory will regard welfare to be normatively relevant. The concept is also at the heart of ethical egoism, the doctrine that one ought to do what's best for oneself.

Welfare figures prominently in theories of value. In fact, some theories of value – those that are “welfarist” – maintain that welfare is the only fundamental intrinsic good. Non-welfarist theories must account for the nature and value of welfare as well, since the job of a complete axiological theory is to discover not only what things are good *simpliciter*, but what things are good *for a subject*. It is precisely the theory of welfare that does this job.

Virtue theory, too, is tied up with the concept of welfare. Many virtues and vices have to do with benefit and harm, and these latter terms are just two of the many idioms of welfare. The virtue of beneficence is even etymologically tied to the former idiom. Prudence, a traditional virtue, is cautious oversight of one's own well-being. To exercise the virtue of charity is to bring goods to – i.e., to improve the welfare of – those in need. The vice of selfishness is excessive concern for one's own welfare. Justice, the virtue, also requires reference to welfare in its analysis.

Justice, the broader concept in moral philosophy, is also plausibly understood in terms of welfare. A truism concerning at least one concept of justice says that people should get the amount they deserve. Amount of what? – welfare. It is welfare that

should be distributed in accordance with desert. We might express the relationship between welfare and justice with the slogan that welfare is the “currency” of justice.

A miscellany of other issues in moral philosophy involve the concept of welfare. The concept of equality is naturally analyzed in terms of welfare: a population manifests equality to the extent that its members are equally well off. There are questions about death, such as, Is death bad for the one dies? There are population puzzles, such as, How many people should there be? One answer is, However many will maximize total welfare; another is, However many will maximize average welfare. Applied ethics contains many questions whose answers presuppose views about what welfare is, such as questions about euthanasia, severely impaired infants, animal rights, world poverty, and human cloning. Any question in ethics that turns, at least in part, on questions of benefit and harm, or on what makes life worth living, involves the concept of welfare.

The concept of welfare plays a central role in the theory of prudence. It is an analytic truth that we are prudentially obligated to maximize our own welfare. Substance is added to this shell of a theory of prudence when a theory of welfare is supplied. Many theories of rationality also require a theory of welfare for completeness. One conception of rationality, for instance, says that an action is rational if and only if it maximizes expected welfare.

Not just philosophers are interested in welfare. Economics is concerned, in part, with evaluating economic arrangements, and one way in which an arrangement can be evaluated is in terms of the amount of welfare it contains. Also, it is a common assumption among economists, and one codified in classical utility theory, that one state

of affairs is better for a person than another if and only if she prefers the former to the latter (Broome 2001). This assumption is in fact a substantive thesis about welfare.

A fundamental question in political science is, What system of government is best? Undoubtedly, one feature that can make one system of government better than another is welfare: in some cases, one system of government is better than another because the citizens of the former system are, on a whole, better off than the citizens in the latter system. Only once we determine what welfare is can we arrive at a complete assessment of systems of government.

Finally, the concept of welfare enjoys currency in ordinary thought. As a measure of how well we are faring, welfare is something we care about long before we enter the philosophy room. When we confront important decisions in our lives – where to go to college, what career to choose, whether to marry, in what part of the country to live – we may consider which of one's alternatives is most in one's interest – that is, which one maximizes one's welfare. Many moral and social emotions, such as envy, compassion, love, and greed, are bound up with the concept of welfare. For instance, it may be that to love someone is, in part, to wish him well.

The idioms of welfare are many. We speak of well-being, of benefit, of doing well, of advantage, of something being in one's interest. We speak of quality of life, of a life well worth living, of the good life. We speak of harm, sacrifice, and disadvantage. Each of these concepts is analytically connected to the concept of welfare. Whenever we make a judgment using one of them, the concept of welfare is involved.



If we are to come to a complete understanding of central issues in the normative ethics of behavior, in axiology, in virtue theory, in the study of justice and equality, in moral psychology, in economics, and in political science, we need a theory of welfare. Indeed, we need one simply to understand what is at the bottom of many claims made in ordinary language, as we conduct the ordinary business of life.

I take welfare to be a kind of *intrinsic* but *relational* value. It is natural to distinguish things that are intrinsically good for a person from things that are extrinsically good for a person. Intuitively, all sorts of things can be at least extrinsically good for a person (e.g., money, health, sunshine, an education, ice-cream). But if they are merely extrinsically good, then their goodness is derivative and contingent. The concept of welfare, by contrast, is the concept of something that is good in itself and underderivatively for a subject. But, as the expression ‘for a subject’ suggests, this is nevertheless a relational kind of value. One and the same thing (such as, perhaps, *Hillary’s being happy at noon*) can be intrinsically good for one subject (in this example, Hillary) and be totally worthless for another subject (e.g., someone who died long before Hillary was born).

## 1.2 Theories of Welfare

Welfare theory is not value theory. A theory of value purports to tell us what things are good in themselves, *full stop*. (Said another way: what things are good in themselves *for the world*.) A theory of welfare, on the other hand, purports to tell us what things are good in themselves *for a subject*. These two areas of moral philosophy are independent: no theory of value commits one to any particular theory of welfare,

and no theory of welfare commits one to any particular theory of value. One could consistently be a nihilist about value – by believing there is no such thing – but be a realist about welfare. In fact, I suppose such a position is not uncommon; many seem to think that intrinsic value is “spooky” or “queer,” whereas welfare, they think, is a perfectly sensible thing to believe in.

Theories of welfare should do several things. They should identify the fundamental source or sources of positive and negative welfare. To illustrate, traditional forms of hedonism maintain that pleasure is the fundamental, intrinsic good for a subject and pain the fundamental, intrinsic evil. I will be assuming that the ontological category of the fundamental bearers of welfare-value – the entities that have the property of being intrinsically good for some subject – are *states of affairs* or *propositions*. If hedonism is true, then states of affairs like *Larry’s experiencing 3 hedons of pleasure at noon* will be intrinsically good for Larry.

I do not assume that the fundamental bearers of welfare value must be these relatively “small” or “atomic” states of affairs, ones attributing a property to a person at a time. It could turn out that the basic welfare value states are much larger, molecular states of affairs – perhaps even entire possible worlds.

Since we are interested in making comparisons between different items – such as certain events or states of affairs, outcomes of certain actions, or entire possible worlds – in terms of their value for some person, a theory of welfare should include quantitative principles. It should identify the factors that determine how intrinsically good an intrinsically good thing would be for someone. A classical form of hedonism, for instance, may claim that the intrinsic value of a pleasure is a function of its duration

and intensity. It might then go on to claim that the intrinsic value of a larger item, such as a person's entire life, is simply the sum of the intrinsic values of all the pleasures and pains within that life. But we should not assume straightaway that the value of a whole life must be "built up" in this way out of the values of its parts. It may be that certain relatively simple states of affairs have certain values, and that certain complex states of affairs built out of them also have a certain values, but that the values of the complex wholes are in no way determined by the values of the smaller parts.

Following Parfit (1984, p. 493), many writers recognize three main kinds of theory of welfare: hedonistic theories, desire-fulfillment theories, and objective list theories (although Parfit also discusses (pp. 501-502) cross-category "composite" theories). According to hedonistic theories, pleasure is the only fundamental, intrinsic good for us. According to desire-fulfillment theories, it is the fulfillment of desire. According to objective list theories, there is a list of things that are intrinsically good for us to have, whether we want them or get pleasure from them.

We have already noted that theories of welfare are not theories of value. They are also not theories about the other ways in which we can rank lives, or rank "how someone is doing." In addition to ranking a life in terms of the amount of welfare in it, we at times have occasion to rank a life in terms of (i) the level of *moral virtue* the person achieved, (ii) the level of *non-moral virtue*, or *excellence*, the person achieved, (iii) the actual *achievements* the person has made, (iv) the *beneficialness* of the life, or how good the life made the lives of others, (v) the *contributory value* of the life, or the amount of intrinsic value the life contributed to the total intrinsic value of the world, (vi) the level of *dignity* of the person, and (vii) the *narrative value* of the life, or how

good the story would be if the life were made into a story. There may be other scales along which we might evaluate a life. All of these scales are interesting and important, but they are not our topic here. And unfortunately, they cloud our topic. If we judge some life to be one that we would never choose, or we judge it to be in some way objectionable or deficient, we must not immediately infer that it ranks low in terms of welfare. Perhaps it ranks low on some of the other scales, and perhaps these scales have no necessary connection to welfare.

Theories of welfare are also not theories about what *causes* human beings to be well-off. Self-help guides often offer this type of advice. They might say that to be truly well off, we need a healthy marriage, a job doing something we love, and fifteen minutes every day in quiet meditation with the accompanying relaxation tapes. I suspect this advice would be compatible with any of the leading theories of welfare, and so it does not answer the question of welfare. The question of welfare is not, What causes us to be well off?, but rather, In what does being well off consist?

It is tempting to think that when the ancients asked about “The Good Life,” they had in mind our topic here. But given the theories some of them proposed, it may be that when ancient philosophers speak of The Good Life, they have in mind a life that ranks reasonably well on *all* (or at least many) of the scales on which lives are ranked. The Good Life may be a life not only with lots of welfare, but with moral and intellectual virtue, with achievement, and with dignity. Perhaps in this context the phrase ‘The Good Life’ means the *most choiceworthy* life. So the final thing that theories of welfare are not: theories of the most choiceworthy life. Welfare provides

one reason to choose a life, but it is plausible to suppose that the others also provide such reasons.

### **1.3 Why Some Are Attracted to a Desire Approach to Welfare**

Roughly, the desire-satisfaction theory of welfare says that one's life is going well to the extent that one is getting what one wants. The desire-satisfaction theory may currently be the most popular view of well-being among philosophers, political scientists, and economists. Aside from the theory's sheer intrinsic appeal – it just sounds plausible to say things are going well for you when you're getting what you want – there are a host of other reasons why the theory is attractive to many:

a. Externalism. Some people think that the thought experiments involving devices like the “experience machine” show that standard hedonism is false (Nozick, 1974, pp. 42-45). Maybe the reason is that they show that how well one's life goes depends upon things other than one's internal mental states – you're better off if you have real experiences and achievements and friends and lovers, say, rather than just imagined ones. We can call the thesis that a person's well-being supervenes on more than just his mental states ‘externalism about welfare’. Standard forms of the desire theory are externalist in that, according to them, your welfare can depend in part on how the external world around you is. If you think externalism is true, then you might find desire satisfactionism attractive.

b. Pluralism. Old time philosophers seemed to have pretty narrow conceptions of the good life. It seems we have to be philosophers contemplating the forms, or perhaps hedonomaniacs spending hours on end in an orgasmatron, to lead ideal lives.



Many people nowadays think this is mistaken. We have more pluralistic leanings: many wildly different sorts of lives can be worth living. Desire satisfactionism can be seen as pluralistic in spirit if monistic in letter, because drastically different lives can be good according to it, so long as it is the life the subject wants (Sumner 1996, pp. 122-23). The desire theory thus seems to provide a way to account for our pluralistic intuitions in a simple, and therefore theoretically attractive, monism.

Don't confuse the point made in the last paragraph with another line of thought. That thought is that we are pluralistic in our valuing. We don't value just one thing, such as, say, pleasure, as hedonists seem to be saying. We value friendship, the welfare of others, freedom, privacy; the list is long. The point of this line of thought is that our theory of value should reflect this. This is all true, and, in my view, it does provide the basis for an argument against a hedonistic theory of *value*. But our concern here is with theories of *welfare*. One can hold a monistic theory of welfare while maintaining a pluralistic theory of value.

c. Internalism. Some philosophers think that only internalism about what is good for one is plausible – that is, that it must be the case that if something is good for a person, she must care about getting it, it must move her, she should be motivated to seek it (Rosati 1995, p. 300; Noggle 1999, p. 303). Some theories of welfare are externalist in the sense that a person could get something that the theory says is good in itself for her without her caring about it (say, a sensation of pleasure, or the performance of a characteristic activity, or the contemplation of a Form). The desire approach seems not to run into this alleged problem; it seems to entail internalism. This



is because the motive to seek what is good for one is built in, since what's good for one, on the theory, is getting what one wants.

The fact that desire satisfactionism can be seen as both externalist *and* internalist is at least an indication of the overuse of these terms. Is it also an indication of a tension within the theory? Maybe so. Part of the idea of internalism (in the sense of c.) seems to be that we must get a kick out of getting what is good for us. But if externalism (in the sense of a.) is true, then it is possible for us to be benefited without even knowing it. There are various ways for desire satisfactionists to rein in their theory, making it more internalist (in the sense of c.). We will touch on some of these ways later on.

d. Liberalism. Many philosophers are liberal minded. They value personal autonomy; they think people should be allowed to do what they want with their lives, to be free to pursue "their own idea of the good life." Some theories of well-being, however, seem rather paternalistic. These theories say: "I know what's good for you, whether you like it or not. It's pleasure, or it's acting in accord with your proper function, or it's the life of philosophical contemplation." Some like the desire-satisfaction theory because they see it as unlike this. "You don't want pleasure?" the theory asks. "Well, that's ok; you can still do well without it. In fact, why don't you just pursue whatever you want. Who are we to say what's good for you?" The thought is that desire satisfactionism is "in tune with the liberal spirit of the modern age, which tends to see human agents as pursuers of autonomously chosen projects" (Sumner 1996, p. 123).

e. Empiricism. Economists like the desire-satisfaction theory because, as empirically-minded scientists, they want a measure of welfare that is also measurable. It is thought that our preferences are observable and hence measurable, even quantifiable, because they are revealed through our choices. I am less sanguine about the measurability of welfare given desire satisfactionism, but that has traditionally been one attractive feature of the theory (see Sumner 1996, ch. 5.1 and Carson 2000, p. 69).

f. Methodological Desire Satisfactionism. In philosophical debates about well-being, it is a common move, when explaining why such-and-such life is or isn't good for the person, to appeal to what the person wants. We can bully our interlocutors into accepting claims assessing the value of a life with the questions, Would you want this life?!, Would you want this life for your child?! This appeal to desires in our methodology may be evidence that desires play a fundamental role in the good life (Carson 2000, p. 69).

g. Sensible Anti-Chauvinism. Platonic Theories and Aristotelian Theories (under one interpretation of Aristotle) imply that only human beings, and in fact only pretty smart human beings, can lead decent lives. The lives of sufficiently unintelligent people, infants and chimps, dogs and cats, are worthless. Another interpretation of Aristotle says that not only can any animal lead a good life, but so can plants, and even artifacts. These extreme views are difficult to accept. Desire Satisfactionism, on the other hand, may strike a nice balance. It "circumscribes the class of welfare subjects in an intuitively plausible way" (Sumner 1996, p. 123): the things that can be well-off are the things that can have desires.

h. Subjectivism/Anti-Realism. Some people think that “objectivism” and/or “realism” in ethics is a pipe-dream. The idea that ethical truths are “out there” for us to grasp with the natural light of reason is repugnant to them. So they might subscribe to “subjectivism” or “anti-realism.” Maybe the idea behind these theories is that the truth about what’s good is determined by our attitudes about what’s good. They see desire satisfactionism as a species of subjectivism, because what is good for someone is determined by his wants.

#### 1.4 Some Metaethical Remarks

Since my project fits squarely within normative ethics, I should think that much of what I say will be compatible with most metaethical views about the status of welfare judgments. Nevertheless, in case it is thought to matter, I will state (but will not here defend) my metaethical convictions. I accept *cognitivism*, and so maintain that welfare judgments (and, incidentally, other normative judgments) have truth values. I accept *realism*, and so maintain, contrary to an error theory, that (simple, positive) welfare judgments are sometimes true. I accept *objectivism* (which is to say that I reject subjectivism), and so maintain that judgments about what would be good for someone are not somehow made true by anyone’s attitude towards the purported good. Such judgments are made true by the way the world is, no matter what anyone thinks about it.<sup>1</sup> I do not have settled views concerning the naturalism/non-naturalism controversy in

---

<sup>1</sup> This taxonomy is confusing because the desire theory of welfare – the theory in normative ethics – is often classified as a form of subjectivism about welfare. This is because desire theories, unlike objective list theories, make welfare depend upon the attitudes the subject has about the objects in his life rather than upon the nature of those

metaethics. Finally, as regards moral epistemology and methodology, I accept and employ something along the lines of the method of reflective equilibrium.

---

objects themselves. So there is a sense in which I *am* a subjectivist: I am a subjectivist not about welfare judgments but about welfare itself.

## CHAPTER 2

### THE PROBLEM OF DEFECTIVE DESIRES

The desire-satisfaction theory of welfare says, roughly, that one's life goes well to the extent that one's desires are satisfied. On standard "actualist" versions of the theory, it doesn't matter what you desire. So long as you are getting what you actually want – whatever it is – things are going well for you.<sup>1</sup> There is widespread agreement that these standard versions are incorrect, because we can desire things that are bad for us – in other words, because there are "defective desires." The aim of this paper is to defend the actualist desire-satisfaction theory against the problem of defective desires. I aim to show how the theory can accommodate the obvious fact that we can desire things that are bad for us. Admittedly, there are kinds of allegedly defective desire the theory cannot accommodate, but these desires, I argue, turn out not to be defective in the relevant way.

#### 2.1 The Problem of Defective Desires

Cases abound in which it is bad for a person to get what he wants. We have *ill-informed desires*. Suppose I'm thirsty and therefore desire to drink from the river. Suppose also that the river is polluted and that drinking from it would make me sick (Carson 2000: 72-73). Clearly, it would be bad for me to drink from the river, even

---

<sup>1</sup> The theory is called 'actualist' to distinguish it from a popular variant: an "idealist" theory, which counts the satisfaction not of one's actual desires but of one's "ideal desires" – typically, the desires one would have if one were fully informed and free from cognitive error.

though it is something I want to do. If only I had more information, I would no longer desire to drink from the river, and would consequently be better off.

We have *irrational desires*. Suppose I need to see a dentist for a procedure that will require drilling. And suppose this time lack of information isn't my problem: I know it is in my long-term interest to see the dentist. But still I want no part of it. Clearly, it would be bad for me not to go, even though it is something I want. If only I weren't weak-willed, I would come to desire to see the dentist, and would consequently be better off (Schwartz 1982, p. 196).

We have *base desires*. Suppose my strongest desires would be satisfied by a perpetual indulgence in bestiality (Moore 1993, pp. 146-147). Some venture to say that any theory that implies that the satisfaction of such desires is good is as false as it is paradoxical. Actualist desire theories are thought to have this implication.

We have *poorly cultivated desires*. Suppose I prefer Muzak over Mozart, mashed potatoes over masterpiece cuisine, and pushpin over poetry. It is thought that I would be better off if I desired the better things, and got them, than if I desired and got the mediocre things I now desire (Schwartz 1982, p. 196).

These last two objections are similar. The idea behind them might be (though need not be) that certain things are somehow inherently more worthy of desire than other things, and this desire-worthiness translates into enhanced welfare when the worthy things are gotten. The objects of base desires are at one end of the spectrum, such that getting them is positively bad; the objects of noble desires are at the other end, and getting them is super; somewhere in the middle are the mediocre things, and getting them isn't so great. That we put effort into cultivating a taste for certain things we think



worthy, and that we encourage some desires over others in our children and other loved ones, suggests that the idea that the objects of desire can vary in worth is common.

Yet another variation of this same idea is found in the possibility of *pointless desires*. Suppose my only desire is “to count blades of grass in various geometrically shaped areas such as park squares and well-trimmed lawns” (Rawls 1971, p. 432). All I want is to count the blades of grass, all day, every day. It is thought that I would be better off if I desired to do something more worthwhile, and did that, than if I did the utterly pointless thing I now desire to do.

Continuing the litany, we have *artificially aroused desires*. Suppose as a result of a late-night indulgence in infomercials, I want nothing more than to own a Flowbee Vacuum Haircut System. Suppose the desire is artificially aroused: I have it only due to the in-your-face marketing campaign’s effect on my insomniatic mind. It is tempting to think that the satisfaction of such desires does not benefit their subject.<sup>2</sup>

Finally, we *desire to be badly off*. Suppose that, guilt-ridden by past crimes, I seek to punish myself by taking an arduous, boring, and insignificant job (Kraut 1994). Suppose I land one and become pretty miserable. Clearly, it is bad for me to get the job, even though it is what I want. If I were less susceptible to guilt, I would have no desire for such a job, and would consequently be better off.<sup>3</sup>

---

<sup>2</sup> I thank Jason Raibley for encouraging me to discuss this class of allegedly defective desire. I cannot here discuss Parfit’s drug-addiction case (1984, p. 497), which might be classified as another kind of case of artificial arousal.

<sup>3</sup> A class of desires I will not discuss here are what Robert Adams calls idealistic desires (1999, p. 88). We desire idealistically when we sacrifice the maximization of our own welfare for something else we value, such as virtue, knowledge, dignity, or the welfare

The problem of defective desires, in one or another of the variants cataloged above, is regarded as a decisive objection to actualist desire-satisfaction theories of welfare by quite a number of writers on the topic. Robert Adams (1999, p. 87) considers it to be “the gravest disadvantage of the theory.” James Griffin (1986, p. 10) maintains that this “objection to the actual-desire account is overwhelming.” Fred Feldman (2004, p. 16) deems it “the central problem” for actualism. Endorsement of the objection is given also by Elizabeth Anderson (1993, p. 129), Richard Brandt (1979, p. 6), Thomas Carson (2000, pp. 80-82, 88-92), Shelly Kagan (1998, pp. 38-39), Richard Kraut (1994), and Thomas Schwartz (1982, pp. 195-197).

## 2.2 The Actual Desire-Satisfaction Theory

What is the theory thought to be so refuted by the problem of defective desires? First, assume that desires are propositional attitudes. *Desire satisfactions*, then, are states of affairs in which a subject desires that some proposition be true and in fact the proposition is true. *Desire frustrations* are states of affairs in which a subject desires that some proposition be true but the proposition is not true.

All parties to the debate seem to agree that any desire-satisfaction theory must count only *basic* (or intrinsic) desires (Brandt 1979, p. 111; Parfit 1984, p. 117). If I desire to turn on my CD player only because I desire to hear the Pixies, I am made better off only if the latter, basic desire is satisfied. My desire to turn on my CD player is merely *instrumental* (or extrinsic): I have it only because I desire something else, and

---

of others (see also Overvold 1980). In Chapter 3, I attempt to show how actualist desire theories of welfare can accommodate these self-sacrificial desires.

I think the thing instrumentally desired will lead to that something else. The theory is therefore interested only in “basic desire satisfaction” – that is, the satisfaction of one’s basic desires. Intuitively, a basic desire is a desire for something for its own sake, not merely for something else that it will lead to, or otherwise realize.

The simplest possible actualist desire-satisfaction theory of welfare would contain the following theses:

- (i) Every basic desire satisfaction is intrinsically good for its subject; every basic desire frustration is intrinsically bad for its subject.
- (ii) The intrinsic value for its subject of a basic desire satisfaction = the intensity of the desire satisfied; the intrinsic value for its subject of a basic desire frustration = – (the intensity of the desire frustrated).
- (iii) The intrinsic value of a life (or segment of a life) for the one who lives it (in other words, the total amount of welfare in the life (or life-segment)) = the sum of the intrinsic values of all the basic desire satisfactions and frustrations contained therein.

We can take this theory to be a paradigmatic version of an actualist desire-satisfaction theory, and the target theory of the problem of defective desires.

The theory is *actualist* rather than *idealist*, since it counts towards welfare the satisfaction of one’s actual desires, not the desires one would ideally have. The theory is *summative* rather than *global*: it implies that the amount of welfare in a life is obtained by summing over all the satisfactions in the life, not just the satisfactions of one’s “global desires” (the desires about one’s whole life, or about a largish part of one’s life).<sup>4</sup> The theory is a *satisfaction* version rather than an *object* version of the

---

<sup>4</sup> On global desires and on the distinction between global and summative theories see Parfit (1986, pp. 496-499) and Carson (2000, pp. 73-74). Though I use Parfit’s terminology here, I think the label ‘summative’ is misleading. Even Parfit’s global theory is summative in the sense that the welfare value of a whole life, according to the

desire theory, since it is desire satisfactions rather than the objects of desire that are intrinsically good and bad for a subject (Rabinowicz and Österberg 1996; Bykvist 1998). Importantly, I mean the theory above to require *concurrence*: in order for a state of affairs to count as a genuine instance of desire satisfaction, the state of affairs desired must obtain *at the same time* that it is desired to obtain. If I desire fame today but get it tomorrow, when I no longer want it, my desire for fame was not satisfied. A desire of mine is satisfied only if I get the thing while I still desire it, and continue to have the desire while I'm getting it. The theory above therefore does not take the *duration* of a desire to be as prudentially significant as some have taken it to be. It doesn't matter, welfare-wise, how long I desire something before I get it. All that matters are the following: that I still desire it while I'm getting it; how long the concurrent desiring and getting last; and how intense the desire is while it is being satisfied.<sup>5</sup>

### 2.3 The Form of The Argument

We can abstract away from the particulars of the types of argument outlined in Section I and be left with the general form of the argument from defective desires:

---

theory, is equal to the sum of the values of the smaller bits of welfare in the life (in this case, the satisfaction or frustration of global desires).

<sup>5</sup> What if the intensity of the desire changes over the time that it is being concurrently satisfied? We can avoid this by defining desires as things that occur at instants (or at very brief intervals of time). At each brief interval at which some concurrent desire satisfaction occurs, there is just one intensity. We can say that the intensity of a desire has 'changed' when a person has a desire of some intensity for some proposition at some brief interval and then, at the next brief interval, has a desire of a different intensity for the same proposition.

- (1) There are defective desires: desires whose satisfaction does not make their subject better off.
- (2) But if the actualist desire-satisfaction theory is true, then there are no defective desires.
- (3) Therefore, the actualist desire-satisfaction theory is not true.

If these premises are true and univocal, actualist forms of the desire-satisfaction theory are done for. Call this the ‘Main Argument’.

Though the Main Argument captures the structure of all the arguments suggested above, the arguments subdivide reasonably well into three categories: (i) arguments from “all-things-considered defective” desires, (ii) arguments from “intrinsically defective” desires; and (iii) a third category to be named later.

## 2.4 Arguments from All-Things-Considered Defective Desires

Some things are intrinsically bad for people. It is the job of a theory of welfare to identify these things. According to hedonists, episodes of pain are intrinsically bad for people. According to actual desire-satisfaction theorists, desire frustrations are intrinsically bad for people. In each case, the allegedly bad items are bad in themselves, rather than for what they lead to. The items would still be bad even if they led to nothing else bad. We can put this more precisely, and theory-neutrally, as follows: a state of affairs *p* is *intrinsically bad* for someone *S* iff given two lives exactly alike except with respect to *p*, the *p*-life is worse for *S* than the not-*p*-life.

Other things are bad for people in a different way. They are bad not in themselves but for what they lead to, or what they prevent. Many desire satisfactions are like this – e.g., many ill-informed desires and irrational desires of the sorts described



above (and likewise for some of the other kinds of desire catalogued above, but more on that later). Recall my ill-informed desire to quench my thirst by drinking from the polluted river. There is nothing intrinsically or necessarily bad about having my thirst quenched. The desire is defective merely accidentally: its satisfaction just happened to be bad overall, in the long run, due to the fact that the river happened to be polluted.

Let's say that if something is bad in this way, then it is *all-things-considered* bad for the subject. That is, a state of affairs  $p$  is all-things-considered bad for someone  $S$  iff the life  $S$  would lead if  $p$  were to obtain is worse for  $S$  than the life  $S$  would lead if  $p$  were not to obtain. That is, iff  $S$  would have been better off had  $p$  not obtained.

Given this simple distinction, it is not difficult to see how actualist desire-satisfaction theories have the resources to deem ill-informed and irrational desires defective. Despite the fact that the theories declare all desire-satisfactions to be intrinsically good for their subject, the theories can nevertheless also declare some desire-satisfactions to be bad – *all-things-considered bad*.<sup>6</sup> Since intrinsic badness and all-things-considered badness can come apart, the theory is consistent with – indeed, it entails – the claim that many desire-satisfactions are all-things-considered bad. To make this more clear, let's take a closer look at the examples.

#### 2.4.1 The Argument from Ill-Informed Desires

Assume my desire to quench my thirst is a basic desire. Assume the object of this desire is the proposition *that I quench my thirst within the next hour*. We all agree

---

<sup>6</sup> This idea is hinted at by Carson (2000: 72-73) and Schwartz (1982: 197-98). I hope to offer a more developed account of this reply.



that this ill-informed desire is defective because, were I to satisfy it, I would be worse off than I would be if I did not satisfy it. The future I would have if I did satisfy this desire would start out pretty good – my thirst would be quenched. But it would quickly turn bad as sickness set in. The future I would have if I chose not to satisfy the desire to quench my thirst would start out not so great – I would be thirsty. But that's about as bad as it would get (at least as far as this case goes), since I would avoid illness. All told, the life I would lead were I not to quench my thirst then and there – call it 'L1' – is better than the life I would lead – call it 'L2' – were I to quench my thirst then and there.

These are the facts. And actualist forms of the desire-satisfaction theory fully accommodate them. The actualist theory entails that L1 is better than L2 because, due to the sickness suffered in L2 (and despite the extra desire satisfaction (the thirst-quenching within the hour) in L2), L1 contains a more favorable balance of satisfaction over frustration. It avoids those desire frustrations that accompany being ill, whose negative value outweighs the positive value of satisfying the desire to be quenched within the hour. Far from being a counterexample to actualist theories, this case of an ill-informed defective desire actually provides confirmation to actualist desire-satisfaction theories of welfare.

So we say that my desire to quench my thirst in this example is all-things-considered defective. If the concept of all-things-considered defectiveness is the one being used in the Main Argument above, then premise (2) is false: according to the actualist desire-satisfaction theory, there are indeed defective desires.

The other concept was that of *intrinsic* defectiveness, the idea of something being bad in itself for a subject, whether or not it happens to lead to anything else bad or prevent anything good. So long as we attend only to basic desires (as any desire-satisfaction theory requires), it is not plausible to say that merely ill-informed desires are intrinsically defective. If I intrinsically desire to quench my thirst right now, and the desire is satisfied, this is intrinsically good for me. Saying this is consistent with saying that it is not in my best interest to quench my thirst, that I ought not to quench my thirst, and that it would be all-things-considered bad for me to quench my thirst. Once we see what is meant – and what is *not* meant – by saying that the satisfaction of the desire is intrinsically good for me, there is no compelling reason to deny it.

And there are compelling reasons not to deny it. Consider what it is to say that my desire to quench my thirst is intrinsically defective. To do this, we need to compare initial segments of two possible lives. First imagine the initial segment of L1, the life in which I do not quench my thirst. This segment – call it ‘S1’ – ends, let’s say, a few minutes after I choose not to quench my thirst. These last few minutes of S1 are spent with a parched throat and a longing for liquid. Now imagine the corresponding initial segment of L2, the life in which I quench my thirst. This segment – call it ‘S2’ – is just like S1 except that I choose to satisfy my desire to be quenched. The last few minutes of S2 are spent with a wet whistle and little longing. Importantly, S2 ends before any pollutant-induced illness sets in.

Now I ask you to compare S1 and S2. I think you will agree that the total level of welfare is higher in S2 than in S1. For S2 has everything S1 has and more: the satisfaction of one extra, intense, basic desire, the existence of some added enjoyment,

the avoidance of an annoying thirst. Actualist desire-satisfaction theories therefore correctly classify S2 as better than S1.

If the Argument from Ill-Informed Desires is interpreted as about intrinsic defectiveness, then, Yes, according to actualist desire-satisfaction theories, the ill-informed desire is not defective. But this is the correct result. Premise (1) of this interpretation of the Main Argument is false.

#### **2.4.2 The Argument from Irrational Desires**

An analogous treatment is available for the version of the Main Argument based on irrational desires. My desire to avoid the pain of drilling at the dentist *is* defective according to actualist desire-satisfaction theories. It is all-things-considered defective. This is because, were the desire to be satisfied, a less favorable balance of desire satisfaction over frustration would result than would result if the desire were not satisfied. This is due to the desire frustration that would ensue from avoiding the dentist (such as that accompanying toothaches, the inability to eat the foods one likes, etc.). Again, it is not plausible to say that the desire to avoid the pain of drilling is intrinsically defective. If I were to die immediately after the time of the dentist appointment no matter what I did, it would clearly be better for me, in terms of welfare, to avoid the dentist and spend my last minutes doing something I really wanted to do.

### 2.4.3 The Argument from Artificially Aroused Desires

Oscar Wilde quipped that the only thing worse than not getting what one wants is getting what one wants.<sup>7</sup> Sidgwick lamented a related point, noting that the objects of our strongest desires often come to us as “Dead Sea apples,” no longer wanted once they are gotten, “mere dust and ashes in the eating” (1907, p. 110). Sidgwick thought that moving to an ideal desire theory, one requiring full information about the objects of desire, could solve the problem, since we could learn in advance whether the objects of our desires would come to us flat (1907, pp. 110-111).

But actualist desire-satisfaction theories can accommodate the Dead-Sea-apple phenomenon. They can do so by building concurrence into the theory, as I have done above. Genuine desire satisfaction is had only when the desire remains once its object is gotten. The concurrence requirement ensures that the getting of Dead Sea apples doesn’t improve welfare, since the very reason the thing is a Dead Sea apple is that the desire for it has vanished.

I claim that, typically, artificially aroused desires are special cases of the Dead-Sea-apple phenomenon. If my desire for the infomercial gizmo is merely an effect of clever marketing, chances are that once I get the device, I’ll no longer want it. A desire-satisfaction theory with concurrence will not entail that the satisfaction of this artificially aroused desire is good for me. Actual desire-satisfaction theories therefore

---

<sup>7</sup> Or at least these words are commonly attributed to him. What he actually said (in print, anyway) comes to the same thing: “In this world there are only two tragedies. One is not getting what one wants, and the other is getting it. The last is much the worst; the last is a real tragedy!” (*Lady Windermere’s Fan*, 1892: Third Act).

are compatible with the idea that satisfying artificially aroused desires is often not good for us. Such desires are often defective because their objects are often Dead Sea apples.

But what if the device arrives in the mail and I am glad to see it; and I remain satisfied with the product? Actualist desire theories will imply that I'm made better off because of it. I say, anti-paternalistically, that is the correct result. A desire's arising through some process we find objectionable is not on its own enough to make the desire itself defective.

It is important to add that there is another way in which artificially aroused desires can be deemed defective by actualist desire-satisfaction theories: such desires may be all-things-considered defective. Suppose I'm happy to see the infomercial item when it arrives, but that, in order to afford it, I broke my budget. In such cases – and no doubt cases are common in which marketing seduces one into buying what one can ill-afford – the satisfaction of the desire would typically be all-things-considered defective. The life I would lead were the desire not satisfied is better than the life I would lead were it satisfied. This is consistent with actualist desire-satisfaction theories.

Maybe there is yet something else behind the Argument from Artificially Aroused Desires. If I acquire a desire by means of some artificial process (such as overzealous marketing, or brainwashing), it is tempting to think that satisfying the desire is not good for me because I don't "really want" the thing, because the desire conflicts with my "true self," because the desire is in some way "inauthentic." If this is the idea behind the Argument from Artificially Aroused Desires, then the argument will not undermine the actual desire-satisfaction theory. For the only sense that can be made of the idea of a desire being "inauthentic," or in conflict with one's "true self," is that



the desire conflicts with many other desires held by the person. And if a desire conflicts with many other desires held by a person (in that satisfying it means frustrating others), the desire is (or could be) all-things-considered defective: the life the subject would lead were the desire not satisfied is worse than the life he would lead were it satisfied.

#### **2.4.4 The Argument from Poorly Cultivated Desires**

The Arguments from Base Desires, Pointless Desires, and Poorly Cultivated Desires are most naturally interpreted as based on the claim that the desires in question are *intrinsically* defective. I treat arguments from intrinsically defective desires below in Section V. However, particularly regarding the Argument from Poorly Cultivated Desires, the claim might be understood as that such desires are merely all-things-considered defective.

But any such argument will fail, for the same sorts of reasons already discussed. Let me briefly illustrate. One reason it might be true that a person has “poorly cultivated desires” is that the person has a set of desires that will lead to an overall lower balance in his life of desire satisfaction over frustration than he would have if he had some other, readily attainable set of desires. Suppose I have frequent, and frequently satisfied, desires for Muzak and mashed potatoes. Suppose my twin has frequent, and frequently satisfied, desires for Mozart and masterpiece cuisine. How can an actualist desire-satisfaction theory imply that my twin’s set of desires is better?

In several ways. My set of desires might be worse because I’m bound to get bored with Muzak and mashed potatoes. Eventually, I’ll no longer want either of them, but, not having cultivated any more interesting desires, I’m able to get no satisfaction.



Having these mediocre desires therefore makes me worse off, but not because satisfying them is no good; rather, it is because the mediocre things are less able to keep me desiring them. Another reason why my twin might be better off with his desires is that, once desires for such things as wonderful as Mozart and masterpiece cuisine are cultivated, the desires for them are more intense than are the desires for mediocre things. Since the desires are going to be satisfied anyway, one gets more out of them, in terms of welfare, the more intense they are. My mediocre desires never light my fire so much, and I'm made worse off for it.<sup>8</sup>

These are perfectly sensible ways in which one can say that one's set of desires is poorly cultivated without having to say that any desire is intrinsically defective. But maybe promoters of the Argument from Poorly Cultivated Desires mean to be saying something stronger: that the satisfaction of an intense desire for Muzak is less good *in itself* for its subject than the satisfaction of an equally intense desire for Mozart. To arguments of this sort I now turn.

## 2.5 Arguments from Intrinsically Defective Desires

Sometimes critics of actualist desire theories have in mind an objection that is different in kind from those so far discussed, despite the fact that we can classify it under the same name. Some critics think that some desires are "intrinsically defective."

---

<sup>8</sup> Note that we're not *always* better off cultivating desires for the refined. It's a contingent matter whether some individual would be better off cultivating such desires. If, say, one will never be able to afford fancy dinners and tickets to the symphony, and so will be stuck with Muzak and mashed potatoes no matter what one desires, one will probably be better off keeping the mediocre desires. I thank Maurice Goldsmith for this point.

The claim is that, for some desires, a person can be made worse off by having them satisfied, not because the satisfaction leads to a lower net balance of satisfactions, but simply because the satisfaction is bad (or not very good) in itself. The case of base desires is perhaps the most common example.

### **2.5.1 The Argument from Base Desires**

In a colorful passage, G.E. Moore presents an argument from base pleasures in an attempt to refute a hedonist theory of value (1993, pp. 146-147). But the point can easily be modified to apply to actualist desire-satisfaction theories of welfare. Modified Moore has us imagine a scenario in which people have no desires for the refined – no desires for the “contemplation of beauty” or for “personal affections” – but in which instead the strongest possible desires are satisfied “by a perpetual indulgence in bestiality” (1993, p. 147). The upshot is that the advocate of an actual desire-satisfaction theory is forced to maintain, absurdly, that “such a state of things would be heaven indeed, and that all human endeavors should be devoted to its realization” (1993, p. 147).

Another provocative case (due to Feldman 2002) provides the basis for a similar argument. Suppose a terrorist hates children and wants nothing more than to see them suffer; suppose he devotes his life to an ongoing terror campaign, and succeeds. None of us is inclined to think the terrorist’s life is worth emulating, but an actual desire view must judge it to be a very good one.

An actualist reply begins by examining the cases more closely. In the Moorean example, a series of claims is made about the desire-satisfaction theory. It is claimed

that (a) the theory entails that the scenario of perpetual indulgence so described would be very good, and in fact much better than the current state of the world. It is also claimed that (b) according to the desire-satisfaction theory, we should see to it that such a scenario is actualized. Finally, it is suggested, though not explicitly stated, that (c) the people in such a scenario would be very well off according to the desire-satisfaction theory.

It is important to distinguish these claims because, in my view, only one of them – (c) – is true. The desire-satisfaction theory, as a theory of *welfare*, itself implies nothing about the value of complex scenarios containing many welfare subjects. Theories of welfare do have implications for how well things are going for each subject in Moore's indulgence scenario. But it is a step to go from the claim that everyone in some scenario is doing well to the claim that the scenario itself is good. The step must be bridged by a principle about how to compute the value of "situations," or possible worlds. But such principles move beyond the scope of theories of welfare and into the jurisdiction of full-blown axiologies. Since the desire-satisfaction theory of welfare contains no such principles, claim (a) is false. An argument based on it therefore won't get off the ground.<sup>9</sup>

Likewise, the desire-satisfaction theory of welfare does not imply that we ought to devote all of our efforts to the actualization of the indulgence scenario. Even if we stipulate that the indulgence scenario is the best scenario available to us, a theory of

---

<sup>9</sup> Let me emphasize that this kind of rebuttal may be ineffective against the argument of the real G.E. Moore, whose target appears to have been a hedonist theory of *value* rather than a hedonist theory of *welfare*.

welfare still does not imply that we ought to actualize it. If theories of welfare have any implications about behavior, they imply only what we *prudentially ought* to do. But since sometimes I prudentially ought to do what I morally ought not to do, and sometimes I morally ought to do what I prudentially ought not to do, theories of welfare on their own imply nothing about what we morally ought to do.<sup>10</sup> So claim (b) above is false, and any argument against an actualist desire-satisfaction theory based on it will be unsound.

Theories of welfare do indeed have implications about how well off a perpetual indulger in bestiality would be. Since it is possible for a person to have strong desires in that direction, and have them satisfied, and satisfied without concomitant frustrations, actualist desire-satisfaction theories of welfare do indeed imply that things can be going very well for a perpetual indulger in bestiality. But I maintain that the actualist desire-satisfaction theories deliver the correct verdict here.

To see why, we need to get very clear on what we're imagining vis-à-vis the thought experiment, and on what we are – and are *not* – saying in claiming that the perpetual indulger has a high level of welfare. For the Argument from Base Desires to work, it would need to establish that the desire (in humans, at least) for sex with animals is *intrinsically* defective. If the claim is merely that such a desire is all-things-considered defective (because, say, the indulger would get bored or lonely, or would come to feel extreme shame or guilt, or would come down with some barnyard

---

<sup>10</sup> Here I am interpreting the word 'should' in the Moorean quotation "all human endeavors should be devoted to its realization" to be the *moral* 'should'. But even if Moore had some other sense of 'should' in mind, such as the *just-plain* 'should', the rebuttal still stands, since the just-plain 'should' and the prudential 'should' likewise come apart.

diseases), then the actual desire-satisfaction theory will agree. To be sure our judgment is not clouded by such considerations, we should imagine a perpetual indulger who would suffer no such maladies. We need also to imagine a person who has genuine desires to have sex with animals, not a person who finds it disgusting but is forced to do it. So it's probably no good to imagine yourself in the barnyard, with all of your current desires. The subject of our Moorean thought experiment must want and get sex with animals, and without large amounts of ensuing desire frustration.

What are we *not* saying if we agree with the desire-satisfaction theory that the perpetual indulger does have a high level of welfare? We are not saying that it is good *simpliciter* that he is doing what he is doing. The desire theory of welfare says nothing about what is good *simpliciter*, only about what is good *for a subject*. It is coherent to say that every indulger in Moore's "heaven" is well off even though it is a very sad state of affairs.

Nor are we giving our approval of the indulger's behavior. It may turn out that what the indulger is doing is morally wrong, and if it is, this is consistent with the desire theory of welfare. Nor are we saying that the life of the perpetual indulger ranks high on the other scales on which we rank lives, such as the scales that measure virtue, dignity, or achievement. In short, it is perfectly consistent for an actual desire-satisfaction theorist to issue the following judgment about the perpetual indulger: "What a pity! Sure, he's well off there in the barnyard, happy doing his thing, getting just what he wants, but his life is pathetic: he will achieve nothing; what he does is degrading; and his moral character is woefully underdeveloped. I would not wish this life upon anyone."



Notice that we have pointed to another way in which a desire can be all-things-considered defective. A desire can be all-things-considered defective not because it adversely affects one's net level of welfare, but because it adversely affects one's character, or one's dignity, or what one achieves in life, each of which corresponds to a scale of evaluation of lives that we care about. Actually, it is not obvious that this sort of defectiveness is merely accidental, at least not for all of these alternative scales on which a life can be measured. For example, perhaps there is a necessary connection between bestiality and low dignity. If so, then we do have a case of an intrinsically defective desire, but not one whose defects contaminate welfare. We therefore need a distinction among types of intrinsically defective desire – there are those that are welfare-defective, virtue-defective, dignity-defective, and achievement-defective. (There may be others for any additional scales on which a life can be ranked.) The lesson is that the Moorean argument must find an intrinsically welfare-defective desire, not merely an intrinsically defective desire or an all-things-considered welfare-defective desire.

We should acknowledge one final factor that might be judgment-clouding: that we find sex with animals disgusting. But that we are viscerally repulsed by some state of affairs does not mean we have good reason to condemn it. Indeed, I think it means we must be very careful not to jump to prejudicial conclusions without giving the matter some thought in a cool moment.

I acknowledge that the disfavor commonly felt towards a life like the perpetual indulger's may survive reflection. But the reason, I submit, that we react strongly against a life of perpetual indulgence in bestiality is that we care about more than just



welfare. We don't merely want ourselves (or those we love) to be well off; we also want (or want them) to do good things, to be good people, to achieve worthwhile goals. A life of perpetual indulgence – in bestiality or whatever – lacks the other elements that we care about.

We have teased out what is and what is not implied by the desire-satisfaction theory. The theory implies that the desire for a perpetual indulgence in bestiality is not intrinsically welfare-defective. That is, it implies that the satisfaction of this desire, on its own, augments its subject's welfare. It is consistent with the desire theory that the satisfaction of this desire is bad *simpliciter*, is not in the subject's long term best interest, and diminishes the subject's virtue and dignity. Once all this is pointed out, the Argument from Base Desires loses its thunder. The perpetual indulger is pleased to be with the animals, is getting what he wants, is happy about it all. Given this and everything else that we have said, the claim that the satisfaction of this desire contributes positively to his welfare does not seem implausible.

Let me return briefly to the case of the terrorist, because it has elements that may not be present in the Moorean thought experiment. For instance, just about everyone agrees that the terrorist is a terrible person and that what he is doing is morally reprehensible. A parallel judgment about the perpetual indulger is far more controversial.

The desire-satisfaction theory of welfare implies that things are going well for the terrorist as he delights in the children's suffering. Again, I submit that this is the correct verdict. But let's remind ourselves what this verdict is not. By granting that the terrorist is well-off, the desire-satisfaction theory is not saying that his being well-off is

good; it is not saying that we ought to let the terrorist do what he is doing; it is not saying that the terrorist's life ranks high on the other scales. The theory is just saying that his level of welfare is high as he continues to get what he wants. But isn't this the correct result? I think one reason we are so outraged at the situation is that the terrorist *is* well off despite that fact that he such a horrible person. We find the injustice of it all maddening. Any theory that delivered the result that the terrorist was not doing well could not explain this injustice (Lemos 1994, pp. 43-44; Goldstein 1989, pp. 269, 271).

I conclude that if the Main Argument is interpreted as being about intrinsically defective desires (that is, intrinsically welfare-defective desires), then premise (1) is false: there is no such animal.

### **2.5.2 The Argument from Pointless Desires**

A desire might strike us as intrinsically defective not because its object is base or depraved, but simply because its object is pointless. The most compelling examples of the claim involve entire lives: an afternoon spent tossing a ball against a wall is quaint; a lifetime spent doing so is pathetic. But actualist desire-satisfaction theories will approve of lives spent doing the most pointless things, so long as the person continues to want to be doing the pointless things.

The argument rests on the idea that some states of affairs are intrinsically pointless. What could this mean? It cannot mean that the states of affairs fail to achieve their point (the purpose for which they were brought about), that they are inadequate means to a worthy end. This is "instrumental pointlessness," and desire theories restricted to basic desires ignore the satisfaction of instrumentally pointless

desires, since they ignore instrumental desires altogether. Nor does 'intrinsically pointless' seem to mean intrinsically bad, as the objects of base desires might be. What the terrorist wants is bad. But what Rawls's grass counter wants doesn't seem *bad*; it just seems ... pointless. Might, then, it mean intrinsically morally neutral? I don't think so. Many activities are neither morally good nor morally bad without being what most would want to call 'pointless'. Consider playing the piano, collecting stamps, or climbing rocks.

We can't say that piano playing, stamp collecting, and rock climbing escape pointlessness because their point is the fun of doing so. Counting the blades of grass and throwing the ball against the wall can have a point in that sense, too. The case in question is of a person who has a *basic* desire to count grass.

Consider yet another hypothesis. These activities are unworthy not because they lack moral goodness (or have moral badness) but because they lack *excellence*. And, it might be claimed, human persons, as beings with dignity and worth, ought to pursue the excellent over the pointless. Although I have some sympathy with this way of thinking, if it is what lies behind the Argument from Pointless Desires, I think the argument fails. For excellence is a scale on which a life can be measured that is distinct from the welfare scale. The grass counter's life lacks engagement in the excellent, and is all the worse for it, but this is not to say that its well-being is damaged. Indeed, it is likely that we could harm the guy (i.e., diminish his well-being) precisely by refusing to let him pursue his heart's desire, despite the lack of excellence of the activity.

So my reply to the Argument from Pointless Desires is similar to my reply to the Argument from Base Desires. Each argument posits intrinsically welfare-defective

desires. I deny that there is any such thing. No way of making sense of the notion of pointless desire yields a kind of desire that is intrinsically welfare-defective. I explain our feeling that the lives described in the arguments are lacking by pointing out that the lives rank low on other scales of evaluation that we care about, scales different from the welfare scale.

At this point one might protest: “Then what is welfare?” Am I employing an idiosyncratic concept of welfare, one made so narrow that no objection can touch it? Is my concept of welfare the same one others have been talking about, the one that is relevant to moral theory?<sup>11</sup> I definitely mean it to be, and I believe it is. I do distinguish a life rich in personal welfare from lives rich in virtue, excellence, dignity, achievement, aesthetic value, etc. But this is standard practice; many writers on well-being do precisely the same thing when characterizing their subject matter for their readers (Griffin 1986, p. 23; Sumner 1996, pp. 20-25; Crisp 2001; Feldman 2004, pp. 8-12). And the concept isn’t so narrow or unconnected to other concepts as to be irrelevant to moral philosophy, or to be such that theories about it are unfalsifiable.

Welfare as it is understood here is relevant in the sense that it still provides *reasons for action*: that some state of affairs would be good for some agent gives that agent a reason to bring about that state of affairs. Nothing I have said in responding to the problem of defective desires contravenes this undeniable thesis. I accept (and believe it would be a mistake to deny) that Moore’s perpetual indulger, Feldman’s terrorist, and Rawls’s grass counter have reason to do what they do. Now, what I have

---

<sup>11</sup> I thank anonymous referees for the *Australasian Journal of Philosophy* for pressing this complaint.

said may be in tension with the idea that welfare, by itself, provides “agent-neutral” reasons – i.e., that the fact that some state of affairs would be good for some subject provides anyone capable of bringing it about with a reason to bring it about. But this seems to me to be exactly the right result. In my view, value *simpliciter* (or value for the world) – rather than mere value for a person – is what provides agent-neutral reasons. That is, that some state of affairs would be good provides anyone capable of bringing it about with a reason to bring it about. Of course, whether a state of affairs is good will depend in part on the level of well-being of the people in it. But, I take it, it is no simple function. It will depend on other things as well, such as, for example, whether the people deserve to enjoy whatever welfare they are enjoying, or whether the people are good people.

The actualist theory defended here is definitely not unfalsifiable. In fact, I myself think it is probably false! But this is due to a different objection.<sup>12</sup> My point here has not been to endorse once and for all the particular actualism presented here. My point rather is to show that the problem of defective desires is, after all, no problem for even very simple forms of actualism (and therefore that it is no problem for actualist desire theories generally, including more sophisticated versions designed to avoid the better objections).

---

<sup>12</sup> What I call the objection from remote desires, the most well-known incarnation of which is Parfit’s case of the stranger on the train (1984, p. 494). Griffin (1986, p. 16-17) and Kagan (1998, p. 37) also present versions of the objection from remote desires.



### 2.5.3 The Argument from Poorly Cultivated Desires

Recall me and my twin. I like Muzak; he prefers Mozart. Let's focus on one hour-long stretch of each of our lives. During that hour I sit in the waiting room of a dentist's office relishing the elevator music. My twin is beside me with Mozart's Piano Concerto No. 23 playing through his headphones. Suppose we have equally strong desires to be hearing our respective music types. Suppose all else is equal. Actual desire-satisfaction theories imply that I am just as well off listening to Muzak as my twin is listening to Mozart.

Once an argument from intrinsically defective desire is presented so starkly, doesn't it lose its force? What reason could one give for the claim that I am worse off than my brother simply because I'm listening to worse music? Even granting the point that Mozart is more worthy of desire than Muzak, why should this have any effect on welfare? My brother and I have equally strong and lasting desires for our respective music. We enjoy hearing our respective choices equally. We're equally content sitting there doing it. Why say I'm not as well off? It may be plausible to say that I'm lacking in aesthetic virtue, or even that I degrade myself by listening to such schlock, but this not to say I'm *harmed* by it, welfare-wise.

It is plausible to say that my desire for Muzak may be all-things-considered defective, as we saw in Section 2.4.4. But we also saw there that the desire-satisfaction theory of welfare accommodates the fact that poorly cultivated desires can be all-things-considered defective. On either interpretation, the Argument from Poorly Cultivated Desires misses the mark.



## 2.6 The Argument from Desires To Be Badly Off

Recall the case: guilt-ridden by past crimes, I desire to be badly off, and so find an arduous, boring, and insignificant job, which makes me miserable. Surely, if ever there is a desire whose satisfaction is not good for its subject, the desire to be badly off is it.

Actualist desire-satisfaction theorists can reply as follows. We admit that the satisfaction of a (basic) desire to be badly off is itself intrinsically good for its subject. But we also point out that such desires are defective, according to desire-satisfaction theories, though in a way we haven't yet described. Perhaps the best way to describe the way in which the desire to be badly off is defective is to say that it is *necessarily all-things-considered defective* – all-things-considered defective because its defectiveness is due to the frustration it will lead to, but necessarily defective because it will lead to the frustrations of necessity.

The example will help to clarify. Here is how the actual desire-satisfaction theorist should describe what goes on in the case. I have a job that is arduous, boring, and insignificant. My daily life is thus jam-packed with desire frustration: to feel bored is, necessarily, to want (intrinsically) to be doing something other than what one is doing. For a person to find a task arduous is also, necessarily, for him to have certain desires frustrated. Does it make sense to say that someone found some stretch of time boring, or arduous, but that he was getting everything he wanted at every moment of that stretch of time?

So each day at this job, I'm miserable. My life is filled with desire frustration. But that's not the whole story, for I do have at least one desire satisfaction on my plate:

my desire that I be badly off. That's one point in my favor against all the points against me. The satisfaction of this desire to be badly off must, of necessity, count for less, in terms of welfare, than all the daily frustrations I rack up. If it were otherwise, then I wouldn't be badly off, and the desire to be badly off would no longer be satisfied.

The sense in which my desire to be badly off is "necessarily all-things-considered defective" is this. Necessarily, any time the desire is satisfied, I will be badly off. This makes it necessarily defective, but it is a form of defectiveness that is accommodated by actual desire-satisfaction theories, as described above. Though the defectiveness is necessary, it is merely all-things-considered (rather than intrinsic); the satisfaction of the desire to be badly off is itself intrinsically good for its subject, because any desire satisfaction is intrinsically good for its subject.

### **2.6.1 The Paradoxes of Desire**

I said that the satisfaction of the above desire to be badly off must, of necessity, count for less, in terms of welfare, than all the daily frustrations I rack up. Otherwise, the desire to be badly off wouldn't be satisfied. But what if the desire to be badly off is stronger than the sum of the frustrations, so that I fail to be badly off?

Let's make the case more explicit. Suppose at  $t_1$  I suffer two desire frustrations, each of intensity 3. It would seem that, given the actual desire-satisfaction theory, my net welfare level at  $t_1$  is  $-6$ . But suppose I also desire at  $t_1$ , to intensity 10, that my welfare level at  $t_1$  be negative. If my welfare level is negative, then my intensity 10 desire is satisfied, which tilts the scales back to the positive side, to  $+4$ . But if that happens, then my intensity 10 desire that my net welfare at  $t_1$  be negative is *not*

satisfied, tilting the scales back to the negative. We have a paradox. My desire that my welfare be negative is satisfied if and only if it is not satisfied.

There is another, simpler paradox in the vicinity. Imagine a person who has exactly one desire: that he be badly off. Our desire-satisfaction theory implies that this desire is satisfied if and only if it is not satisfied.<sup>13</sup>

But not just desire-satisfaction theorists are mired in paradox. Analogous paradoxes get off the ground without assuming a desire theory of welfare. Imagine a person who desires, to intensity 10, that his net balance of desire satisfaction over frustration at some time be negative. Suppose he gets 6 units of desire frustration at that time. It would seem his balance is  $-6$ . But if it is, then his intensity 10 desire is satisfied, and so his net balance is  $+4$ . But then his intensity 10 desire is *not* satisfied. In short, this desire is satisfied if and only if it is not satisfied.

A simpler paradox also exists. Imagine a person who has exactly one desire: that none of his desires is satisfied. This desire is satisfied if and only if it is not satisfied. These paradoxes of desire are everyone's paradoxes.

My response to the paradoxes for desire-satisfaction theories is that they are inherited from these more basic paradoxes. The paradoxes for desire-satisfaction theories are just variants of the paradoxes of desire. One possible answer to the paradox (at least to the simpler version) is to reject the possibility of having just one desire. If the concept of desire is analytically tied up with counterfactuals involving belief and action, then it is probably impossible to have just one desire. An alternative response

---

<sup>13</sup> The paradoxes for desire theories of welfare were brought to my attention by Fred Feldman and by Kris McDaniel. Feldman raises the paradox in print in Feldman (2004, p. 17).

(to the complex as well as the simpler version) is that although it seems for all the world that a person could desire that none of his desires be satisfied, the paradox itself shows that this is impossible (cf. some commonly accepted responses to Russell's paradox and to the liar paradox). It is beyond my scope here to attempt to solve the paradoxes of desire. I will say only that however the more basic paradoxes of desire are solved so will the paradoxes for desire-satisfaction theories be solved.

The paradoxes of desire notwithstanding, the actualist desire-satisfaction theory of welfare stands up well to what is commonly thought to be a devastating objection: the problem of defective desires.

### CHAPTER 3

## DESIRE-BASED THEORIES OF WELFARE AND THE POSSIBILITY OF SELF-SACRIFICE

There is a well-known and widely-accepted argument, due to Mark Carl Overvold (1980), against standard desire-based theories of welfare. The idea is that these theories have the absurd implication that self-sacrifice is conceptually impossible. An act of self-sacrifice would seem to be one in which the agent, for the sake of something or someone he cares about, voluntarily brings about some outcome that he knows is not in his best interest. Standard desire-based theories, Overvold says, identify an agent's best interest with "what the agent most wants to do" (p. 105). So it would seem that, according to the theory, the outcome that an agent wants most is the outcome that is most in the agent's self-interest. Since any ostensible act of self-sacrifice is performed voluntarily and knowingly, its outcome is the one the agent wants most; so it is the act that is most in the agent's interest, according to the theory. But if the act is the one most in the agent's interest, then nothing is being sacrificed after all. Self-sacrifice is therefore impossible, given a standard desire-based theory of welfare.

Many others besides Overvold regard this argument from self-sacrifice to be a successful refutation of standard desire-based theories. Overvold's argument is accepted, for example, by Richard Brandt (1991), Thomas Carson (1993; 2000, pp. 74-76), James Griffin (1986, p. 316, 25n), Brad Hooker (1991), and Wayne Sumner (1996, pp. 134-135). The same kind of argument is made independently by Thomas Schwartz

(1982, pp. 199-200) and by Amartya Sen (1982, ch. 4). A similar line of attack is pursued more recently by Robert Adams (1999, pp. 87-89).

Overvold suggests (1980, p. 118, 10n) a way to revise the theory so as to avoid the offensive implication. For reasons I will not discuss here, I do not think Overvold's new theory is acceptable. But desire theorists need not worry because, as I hope to show, the original argument from self-sacrifice does not threaten the desire-theoretic approach. The most simple, unrestricted desire-based theory of welfare straightforwardly accommodates the phenomenon of self-sacrifice – so long as it is formulated properly.

Crucial is a fundamental distinction between the ways in which a desire-based theory can rank lives (or parts of lives) in terms of welfare. One kind of desire-based theory ranks the possible lives a person might lead by looking to the person's preferences with respect to those possible lives as wholes. The best life for the person is the one the person wants most. This kind of view appears to be the target of Overvold's argument.<sup>1</sup> But according to another (and, I think, rather standard) kind of desire-based, we rank the possible lives a person might lead by looking to how well each of those lives satisfies the desires the person will have if she leads that life. The best life is the one that best satisfies the desires it contains. This latter kind of theory, we will see, is perfectly compatible with self-sacrifice.

---

<sup>1</sup> Admittedly, I am not *certain* this is Overvold's target. But it is at least a view that Overvold's argument appears to succeed at undermining. If Overvold's target is instead the kind of view I mean to be defending here, then one can take this paper as attempting to show that Overvold's argument fails on its own terms.



### 3.1 Preliminaries: Theories of Welfare; The Notion of Self-Sacrifice

The main thing we ask of a theory of welfare is the identification of the fundamental source or sources of positive and negative welfare. Hedonism's answer, to illustrate, is simple: pleasure is the only fundamental source of positive welfare and pain is the only fundamental source of negative welfare. Another thing we ask from a theory of welfare is a ranking of events or states of affairs in terms of how good they would be for some particular subject. We want, for example, a theory of welfare to rank the possible lives a person might lead in terms of how good each would be for that person. Parts of lives are also of interest, as when we consider how good some year was for some person. A "summative" hedonistic theory has a simple answer to this kind of question too: one possible life (or year, or other life-segment) is better for the subject living it than another just in case it contains a greater balance of pleasure over pain for him.

Self-sacrifice has to do with *actions*; actions are the sorts of things that can be self-sacrificial. But theories of welfare themselves say nothing about actions, so some principles connecting the notions of welfare and self-sacrifice are needed. To begin, we should distinguish between putative kinds of self-sacrifice. Two kinds of sacrifice can immediately be put aside. One such kind occurs when a person forgoes some good *unknowingly* and/or *accidentally*. We do this all the time, and it is not what Overvold has in mind. Another kind of sacrifice, if we can call it that, occurs when someone sacrifices a lesser good for himself for a greater good for himself. Such acts might feel sacrificial if they involve delayed gratification, but, in such cases, there is no *net* sacrifice. The kind of self-sacrifice in which Overvold is interested obeys the following

principle: an act exemplifies it only if performing the act makes the agent worse off than he otherwise could have been. In other words, an act is an act of self-sacrifice, in the intended sense, only if some alternative to the act would bring about more welfare for the agent than it. Let's say that an act is *in the agent's best interest* just in case the life the agent would lead were he to perform it is at least as good for him as the life he would lead were he to perform any alternative to it. We can then state the principle relevant to our purposes here as follows:

**A Principle about Welfare and Self-Sacrifice:** An act is an act of self-sacrifice only if the act fails to be in the agent's best interest.

I said above that another necessary condition is that the agent *know* that the act fails to be in his best interest. We can keep this additional condition in mind, but I will stick with the simpler principle above; it's strong enough to do the work we need it to do.

### 3.2 One Kind of Desire-Based Theory of Welfare and The Argument from Self-Sacrifice

In contrast to the summative approach to the ranking of lives (illustrated above) stands one kind of desire theory of welfare. According to the crudest version of this theory, the possible lives a person might lead are ranked directly on the basis of the person's desires about them: one life is better for someone than another just in case he prefers it to the other. Call this theory *Simple Preferentism*.<sup>2</sup>

---

<sup>2</sup> I will assume that preference is reducible to desire as follows: S prefers p to q just in case S's desire that p is stronger than S's desires that q. If a person is averse to some state of affairs, we can say that she desires it to a negative degree. If a person is indifferent to a state of affairs, we can say that she desires it to degree zero. Given this reduction, there is no substantive difference between desire theories of welfare and

The target (as I see it) of Overvold's argument from self-sacrifice is a souped-up version of this subjectivist approach. Simple Preferentism has us look to the *actual* desires a person has concerning the various possible lives she could lead. The one that is most in her self-interest is the one she most desires to lead. According to the souped-up theory, by contrast, we look not at the actual desires the person has for each of her various possible lives but instead at her "rational desires": roughly, the desires she would have if she "had been fully exposed to available information" and were to keep this information "firmly and vividly in mind."<sup>3</sup> To discover the rational desires of some subject, go to the nearest possible world in which that subject is fully informed and vividly appreciates the information (henceforth "fully and vividly informed"). The desires the subject has in this counterfactual situation are the rational desires she actually has.

So Overvold's target appears to be a theory like (or at least one that entails) the following:

**Rational Preferentism:** One life is better for a subject than another iff the subject rationally prefers it to the other.

Incidentally, Rational Preferentism, or something very much like it, seems to be the theory Sidgwick discusses in the following famous passage (1907, pp. 111-112):

... a man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realised in imagination at the present point of time.

---

"preferentist" or "preferentialist" theories of welfare.

<sup>3</sup> Brandt (1972, p. 682), quoted by Overvold (1980) on pp. 106-107. Throw in any of the other ways to idealize desires that have been suggested, such as, for example, full imaginative acquaintance. Nothing in this paper hangs on it.

Rational Preferentism doesn't sound immediately implausible. The other day my spouse and I were trying to choose from among several possible flights overseas. We wanted to pick the best flight – the one that was most in our interest.<sup>4</sup> We did our best to get as much information as we could about each flight – its price, the time it left, the length of any layovers, the reputation of the airline, whether it earned us frequent flier miles. We gathered all the information, kept each factor in mind, and decided which package we most preferred. This common practice seems to be an unconscious endorsement of something like Rational Preferentism. And Rational Preferentism avoids the obvious defect of Simple Preferentism: that whatever we happen to prefer most is automatically best for us. Since we might be mistaken, for example, about the actual length of the layover, we might prefer a flight that is actually worse for us.

Rational Preferentism respects and explains this obvious fact.

But Rational Preferentism, Overvold points out, has an unacceptable consequence. It implies that self-sacrifice is conceptually incoherent. Overvold (p. 108) invites us to

Suppose a man wants more than anything else that his four sons attend a very expensive private college. He is not a rich man. The closest thing to a tangible asset he possesses is a huge life insurance policy. After carefully considering his options, he resolves to kill himself, making it look like an accident. He does so, and four years later his eldest son begins college. Eventually all four sons complete their education and enjoy very happy and rewarding lives.

---

<sup>4</sup> That is, the one such that were we to choose it, the lives we would go on to lead would be at least as good as any of the lives we would lead were we to choose one of the other flights. Ignore the complication that possibly no single flight is best for *both* of us.

Allow me to fill in some additional details to the thought experiment. Let's suppose that the father had just two possible lives open to him. In one – call it 'L1' – the father commits suicide, his sons collect on the insurance, and they go on to "complete their education and enjoy very happy and rewarding lives." The father knows that this is what would happen if he were to commit suicide, and he fully and vividly appreciates these facts. In the alternative possible future, which we'll call 'L2', he does not commit suicide. Let's suppose that he struggles to fund his sons' educations, but is unable to do so. His sons go on to lead decent enough lives, though, we will suppose, less happy and rewarding than if they had gone to college. Let's suppose that in L2 our protagonist would have suffered from periodic spells of guilt at his failure to send his sons to school, but that he would have learned to overcome these feelings. He would have gone on to lead a long and overall reasonably satisfying life, a life well worth living by all accounts, if he were to decide against suicide. We must also suppose that, at the time of his decision, the man knows that this is how things would go if he were to decide against suicide, and, further, that he keeps all of this information vividly before his mind when choosing. We are therefore supposing that his actual preferences (at least concerning this decision) are his rational preferences.

The fully and vividly informed father surveys his two possible lives: the foreshortened L1 where his children thrive and the long and fairly happy L2 where his children do a little less well. He prefers and chooses L1. It is obvious that the correct way to describe this is as an act of self-sacrifice. The man sacrifices a long and decent life for himself and takes instead a life cut short, all for the sake of his children's welfare. Call this case *Sacrificial Father*.



But Rational Preferentism disagrees. Since the father rationally preferred L1 to its only alternative, Rational Preferentism implies that L1 is best for the father, and therefore that the act that brings about L1 is in the father's best interest. But, as The Principle about Welfare and Self-Sacrifice states above, if an act is in the agent's best interest it cannot be an act of self-sacrifice. Nothing is being sacrificed in such a case. Therefore, Rational Preferentism implies that the man's act of suicide is not self-sacrificial. But of course it is: by committing suicide, the man sacrifices a good life for himself in order to send his kids to college. Clearly, the man fares better overall in L2.<sup>5</sup>

### 3.3 Another Kind of Desire-Based Theory of Welfare

What Simple Preferentism and Rational Preferentism (and, incidentally, Overvold's own theory) have in common is the manner in which they rank lives in terms of welfare. Metaphorically speaking, they lay out before the subject the whole lives the subject might lead; then they ask the *subject* to rank the lives: Which life do *you*, dear subject, like best? This is subjectivism *par excellence*. It also seems to be precisely the feature that makes the theory go wrong with respect to self-sacrifice. The approach I favor looks not for which whole life the subject has the strongest desire; rather, it looks for which life is such that all of the desires the subject would have

---

<sup>5</sup> The bulk of Overvold's paper is spent arguing for the stronger claim that, given such a theory, the concept of self-sacrifice is unsatisfiable. This requires an analysis of self-sacrifice, which Overvold provides. I will skip discussing the precise nature of self-sacrifice. Our intuitive grasp of the notion, plus the clarifications made in Section 1 above, will be enough for the purposes of this paper.



throughout that life would be best satisfied.<sup>6</sup> Don't, theory of welfare, ask the subject which life she prefers. Instead, theory, examine the lives yourself. Take note of each occasion in each life in which the subject gets what she wants and each occasion in which she doesn't get what she wants. Whichever life contains the greatest balance of desire satisfaction over frustration is the best life. Never mind which life the subject happens to prefer right now. Since this approach, in ranking lives, de-emphasizes how the subject feels about them, it has a more *objectivist* flavor.<sup>7</sup>

The approach I am rejecting is not only radically subjectivist but also *holistic*: the welfare value of a life is not reducible to the welfare values of its parts. By contrast, according to *atomism* about welfare, there are minimal bits of welfare (the "atoms of welfare"); there are good bits and bad bits; these bits accrue over time in a life; the amount of welfare in a life is determined roughly by computing the balance of good bits over bad bits. The summative hedonism alluded to in the beginning of the paper is both atomistic and objectivist: atomistic because, according to the theory, there are atoms of welfare (pleasures and pains) that accumulate; and objectivist because it ranks a subject's possible lives without consulting the subject to see how she feels about the lives.

Although desire-based theories of welfare are often associated with subjectivism and holism, I offer an objectivist (or at least "objectivistic") and atomistic desire-based

---

<sup>6</sup> cf. Phillip Bricker (1980, p. 384): "In deciding which is the prudent act, the agent must be able to compare, for each of the acts he might perform, the preferences he would have as a result, and the extent to which those preferences would be satisfied."

<sup>7</sup> I think the distinction outlined here bears some similarity to the distinction between "object versions" and "satisfaction versions" of the desire theory in Krister Bykvist (2002, p. 475, 2n).

theory of welfare. The theory I will present is probably the simplest possible such theory – “square one” for atomistic desire theories of welfare. I do not claim that it stands in no need of refinement. My point rather is to show that the even the simplest imaginable form of the desire theory – so long as it is atomistic rather than holistic – easily accommodates the phenomenon of self-sacrifice.

Derek Parfit (1984, p. 494) talks about “Desire-Fulfillment Theories,” the simplest variant of which “claims that what is best for someone is what would best fulfill *all* of his desires, throughout his life.” Richard Kraut (1994, p. 164) talks about views that “equate the human good with the satisfaction of desire”:

Roughly speaking, what makes a state of affairs good for someone is its satisfaction of one of that person’s desires; accordingly, our lives go well to the extent that our desires, or the ones to which we give the greatest weight, are satisfied.

Shelly Kagan (1998, p. 36) talks about a “desire-based or preference theory of well-being”:

According to the preference theory, well-being consists in having one’s preferences satisfied. To the extent that your preferences or desires are satisfied, you are better off; to the extent that your preferences or desires are not satisfied, you are less well off. (Of course not all of your preferences are equally important to you, so we would probably want to make the view more complicated, holding that the stronger the desire, the greater the contribution to well-being made by the satisfaction of that desire).

Each of these sketches of a desire-based theory of welfare can be seen as pointing toward the sort of theory I want to elucidate. When Kraut’s description tells us that “what makes a state of affairs good for someone is its satisfaction of one of that person’s desires,” we might understand this as specifying what the “atoms of welfare” are. When Kagan’s description tells us that “To the extent that your preferences or

desires are satisfied, you are better off; to the extent that your preferences or desires are not satisfied, you are less well off,” we might understand this as suggesting a summative and objectivist approach to the ranking of lives over a holistic and subjectivist approach. Let us pursue these suggestions.

We begin with the atoms: *instances of desire satisfaction*, or *desire satisfactions*. A desire satisfaction is a state of affairs in which a subject desires that some proposition is true and in fact the proposition is true. There are also bad atoms, desire frustrations, which are states of affairs in which a subject desires that some proposition is true but the proposition is not true. It is no part of the theory that a subject needs to experience feelings of satisfaction in order to have a desire satisfied. I assume that the objects of desire are “eternal propositions” – entities whose truth value cannot vary over time.

The first part of the theory says that

- (i) Every desire satisfaction is intrinsically good for its subject; every desire frustration is intrinsically bad for its subject.<sup>8</sup>

This clause answers that question, “What is The Good of Man?” It will be easiest to work up to an account of how these atoms add up to valuable wholes if we first answer a second question: How good are the things that are The Good of Man? What factors determine the precise value of an atom?

Kagan offers the common suggestion that “the stronger the desire, the greater the contribution to well-being made by the satisfaction of that desire.” Following this

---

<sup>8</sup> Substitute ‘non-instrumentally’ for ‘intrinsically’ if you think it matters. The idea of something being intrinsically good *for a subject* is to be contrasted with the idea of something being intrinsically good *for the world*. As I see it, theories of welfare should remain silent about the latter notion. So far as the theory of welfare is concerned, states of affairs that are good for a subject may make the world no better, or may even make it worse.

suggestion, I assume that whenever a subject desires some proposition, there is some degree to which he desires the proposition. This is the strength or intensity of the desire. It is probably not plausible to suppose that every desire has some felt intensity. At noon yesterday, I took a break and went for a walk. This is something I strongly wanted to do – I wanted to do it more than I wanted to do quite a lot of other things I could have done at the time. Nevertheless, there was no phenomenological feel to the desire. So it seems more plausible to suppose that a desire's intensity is closely related to its place in the subject's overall preference ranking. We can get an idea of how strongly a subject desires the propositions he desires by discovering his *preferences*: for each pair of such propositions, we discover which of the pair he prefers. As a rough-and-ready but highly imperfect heuristic device, we can pretend that the intensity of a desire is revealed by how much money the subject would be willing to spend to have the desire satisfied.

Whenever a subject desires some proposition, not only is there some intensity to the desire, there is also some stretch of time during which he desires the proposition. So we have two factors: the desire's intensity and its duration.<sup>9</sup> These two quantities determine a larger quantity, which we can call, for lack of a better term, the “amount of desire” in the desire. We can look at this quantity as representing how important the desired proposition is to the subject. The longer the desire is held and the more intense it is while it is held, the “more desire” there is in it and the more important the desired proposition is to the subject.

---

<sup>9</sup> Since the intensity of a single desire can change over its duration, the real factor is the *average* intensity of the desire over its duration. I ignore this complication in what follows.

All this translates into features of desire satisfactions. For any desire satisfaction (any state of affairs in which a subject is getting what she wants), there is an intensity and a duration. The intensity of the desire satisfaction, we can say, is just the intensity of the desire satisfied. And given the assumption above about eternal propositions, the duration of the satisfaction is likewise just the duration of the desire satisfied. We can then introduce the idea of the “amount of desire satisfaction” in a desire satisfaction, which, as above, is a function of the satisfaction’s intensity and duration.

The simplest kind of atomistic desire theory would have it that the *value* (for the subject) of a desire satisfaction is just the amount of desire satisfaction in it. For any desire satisfaction, the stronger it is and the longer it lasts, the better it is for the subject. Officially:

- (ii) The value for a subject of a desire satisfaction is equal to the amount of desire satisfaction in it; the value for a subject of a desire frustration is equal to  $-(\text{the amount of desire frustration in it})$ .

This is fairly intuitive. The greater the amount of desire satisfaction in a satisfaction, the more important the desire was to the subject; so it makes sense that satisfying the desire should be more valuable to the subject.

Consider some examples. Suppose at 3:00 p.m. Bruce desires to go for an hour-long walk. That is, he desires, we can suppose, somewhat artificially, the following proposition: *that Bruce takes a walk between 3:00 p.m. and 4:00 p.m.* Suppose Bruce takes that walk. And suppose that as he is taking the walk, he continues to be wanting to be taking it. In such a case, Bruce gets an hour-long desire satisfaction of a certain intensity. (Actually, he would probably get more than that, since he probably desired to



take the walk at least a few minutes before taking it. In fact, assuming it is possible to have desires about the past, Bruce could actually get quite a bit more desire satisfaction out of this, so long as he were to continue to desire after the fact that he took the walk.) Consider another example. Suppose I desire in April that the Red Sox win the World Series in October. Suppose in fact the Red Sox will win (quite an exotic thought experiment, I realize). The present suggestion about computing the value of desire satisfactions implies that, as I continue through spring and summer to desire victory for the Red Sox, I am “racking up” welfare (since, I will assume, it is true in the spring and summer that the Red Sox win in the fall). We can look at it like this: a Red Sox victory is better for me if I have desired it for longer because I have more invested in the prospect.<sup>10</sup>

Now that we have identified the bits of welfare and the value of any given bit, we can provide a simple answer to the ranking problem: add 'em up.<sup>11</sup> To determine the amount of welfare in a life, locate all the desire satisfactions and frustrations in the life, determine the value of each for the subject, add up all these values. That is to say:

---

<sup>10</sup> The present suggestion admittedly has a bizarre consequence. Suppose from Monday through Thursday Sally wants to go bowling on Friday night. Suppose on Friday afternoon, Sally changes her mind and prefers to go to the disco on Friday night instead. The present theory recommends that Sally go bowling, since that will issue in a longer (and hence more valuable) desire satisfaction for her. But surely Sally ought to do what she prefers *then*. Various solutions to problems of changing desires have been proposed, but I cannot take up the issue here. For my goal is just to show how the simplest imaginable desire theory – so long as it is atomistic – accommodates self-sacrifice in a straightforward and intuitive way.

<sup>11</sup> *Pace* David Velleman (1991), who argues that the welfare value in a life is *not* just the sum of the welfare values of its momentary stages.

- (iii) The intrinsic value of a life (or segment of a life) for the one who lives it = the sum of the intrinsic values of all the desire satisfactions and frustrations that occur in that life.

Call this theory, the one consisting of (i), (ii), and (iii), *Simple Desire Satisfactionism*.

Clause (iii) entails a principle in direct competition with the formulation of Rational Preferentism above: one life is better for someone than another iff it contains a greater balance of desire satisfaction over frustration than the other.

The theory *is* subjectivist in one sense: according to it, welfare is a matter of the subjective attitudes of the subject of welfare. But the theory is objectivist with respect to the evaluation of whole lives – the subject's attitudes about her life as a whole do not directly determine how good it is. They can, however, *indirectly* determine how good it is. For a subject can have desires about her life as a whole. She may desire, after surveying several possible futures, that she lead *this* one, while having no desires to lead any of the others. If she so desires, and if in fact she leads that one, then her desire is satisfied, and Simple Desire Satisfactionism implies she is thereby better off. So although the theory is atomistic, it accommodates the idea that holistic features of one's life can be relevant to how good it is.

### 3.4 Simple Desire Satisfactionism and the Possibility of Self-Sacrifice

Simple Desire Satisfactionism straightforwardly handles the phenomenon of self-sacrifice. Given Simple Desire Satisfactionism, an act is in the agent's best interest iff no alternative to it would produce a greater balance of desire satisfaction over frustration for the agent. An act can be in the agent's best interest if it leads to the satisfaction of a large number of desires, or to the satisfaction of intense desires, or to

the satisfaction of long-lasting desires, or some combination thereof. Or it can do so if its alternatives would lead to enough frustration, or fail to lead to much satisfaction.

The sacrificial father faces a choice between L1, the suicide life, and L2, the longer life. We can suppose that L1 contains a few desire frustrations in its last moments – those that naturally accompany impending death. L1 also contains some desire satisfactions not present in L2, most notably the desire that the sons attend college. The relevant features of L2 are far too numerous to state, since the life is long. But recall the general course of the life: the man's frustrated desire that his sons attend college continues, but eventually he "gets over it" and goes on to lead a fairly decent life.

Given these descriptions of L1 and L2, it is clear that, according to Simple Desire Satisfactionism, nothing bars L2 from having a far greater value for the father than L1 – this despite the fact that the father prefers L1 to L2. L2 is so much longer that we are free to describe it as having any number of intense and long-lasting desire satisfactions. True, L2 will contain recurring frustrations over the failure of the sons to attend college, but there is nothing to stop us from stipulating that these frustrations are heavily outweighed. So let L2 be such that its balance of desire satisfaction over frustration is far greater than the overall balance in L1. According to Simple Desire Satisfactionism, then, L2 is better for the man than L1. The theory thus opens the door for self-sacrifice, for, according to the theory, a person, even a fully and vividly informed person, can choose what is worse for him for the sake of someone or something he cares about.

### 3.5 Objections and Replies

Or can he? Perhaps Overvold could attempt to stop the train at just this juncture, perhaps by appeal to a principle like the following:

**A Principle about Preference:** Necessarily, if a subject knows that some state of affairs *p* contains a greater balance of desire satisfaction over frustration for the subject than does an alternative state of affairs *q*, then the subject prefers *p* to *q*.

The idea behind A Principle about Preference is attractive enough. Suppose I have complete knowledge about two alternative lives of mine, *S* and *F*. *S* satisfies many of its desires and frustrates none, while *F* frustrates many and satisfies none. How could I possibly prefer *F* to *S*? What features would attract me to *F*? None, it would seem, for if there were any, then *F* would have to satisfy my desires for them.

Despite its initial plausibility, A Principle about Preference does not stand up to scrutiny. The root of the problem is that it is possible for a person to know that he will desire certain things in the future, but yet not be moved in the present to behave in such a way that those future desires will be satisfied. I might prefer today the alternative of not visiting the dentist even though I know that visiting the dentist would better satisfy my desires in the long run. Call me irrational, call me weak-willed – it doesn't change the fact. And besides, not only the weak-willed falsify the principle. Consider an evil genius who gives a good Samaritan two options:

Option A:      a life of decadence with desires for decadence; misery for everyone else without desires to the contrary

Option B:      a life of toil with desires to the contrary; flourishing for everyone else without desires that others flourish.

The good Samaritan sees that Option A would give herself a far greater balance of satisfaction over frustration in her life. But being a selfless saint, she cannot bring herself to choose it. She prefers Option B. So A Principle about Preference is false.

Consider a second objection to my attempt to demonstrate the desire theory's compatibility with self-sacrifice. Perhaps Simple Desire Satisfactionism suffers the fate Rational Preferentism suffers in Sacrificial Father: perhaps even Simple Desire Satisfactionism implies that the father's act is not one of self-sacrifice. I said that a desire's intensity corresponds to its place in the subject's overall preference ranking. Overvold says that the father wants his sons to attend the private college "more than anything else." Perhaps we should interpret this as meaning that the importance of this desire outranks the total importance of all the desires the father would satisfy in the longer life, put together. If this were the case then, according to Simple Desire Satisfactionism, the value of the satisfaction of this super-intense desire would outweigh the sum total of the values of all the desire satisfactions in L2, the longer life. Then Simple Desire Satisfactionism implies that the father does not sacrifice his own interest in choosing L1 after all. So it has not been shown, the objection concludes, that Simple Desire Satisfactionism is compatible with self-sacrifice.<sup>12</sup>

I think this objection can be answered if we divide Sacrificial Father in two. As Overvold presents (and as we embellished) the original case, it is not said whether the father's desire that his sons attend the college is stronger than all the future desires in L2 combined. So let's introduce two new cases, two versions of Sacrificial Father:

---

<sup>12</sup> I thank an anonymous referee for the *Canadian Journal of Philosophy* for raising this objection.



*Fanatical Father* and *Devoted Dad*. In *Fanatical Father*, the fanatical father is as the objection imagines him: the desire that his sons attend the college is stronger than all the future desires combined in his L2. In *Devoted Dad*, the “merely” devoted dad isn’t so fanatical: while his strongest desire at the time in question is that his sons attend the college, this desire is *not* stronger than all the future desires combined in his L2. But the desire is still stronger than his desire not to commit suicide, so even the merely devoted dad commits suicide for the sake of his sons.

I think Simple Desire Satisfactionism generates the correct result in both cases. Take the second case first, which is how I understood Overvold’s original case. As illustrated above, Simple Desire Satisfactionism implies that the dad in *Devoted Dad* sacrifices himself; therefore the theory is compatible with self-sacrifice. We need not include the supposition (present in *Fanatical Father*) that the dad’s present desire outweigh all the future desires combined in the life he doesn’t choose, in order to get him not to choose that life. That is not how desires motivate. A known future desire motivates one presently only if one’s knowledge of it causes a new present desire. But this need not happen in *Dedicated Dad*. As the case is set up, the dad’s situation is like this: he knows L2 would contain a greater balance of desire satisfaction over frustration for him than L1; if his knowledge of those future desires were to cause corresponding present desires, his motivation might tilt in favor of L2; but this knowledge doesn’t, as a matter of fact, cause these desires; so the father continues to prefer L1. As we observed in the previous objection, it is possible for a person to have desires in the future – desires he knows about – that nevertheless fail to move him today.

Consideration of Dedicated Dad shows, then, that Simple Desire Satisfactionism is compatible with self-sacrifice after all.

The objection has therefore been answered. We need cite only one case in which there is self-sacrifice according to Simple Desire Satisfactionism to prove compatibility between the two. Nevertheless, let's consider Fanatical Father. It leads naturally to a third objection. In this unusual variant, the fanatical father's present desire that his sons attend the college is unbelievably intense. The father is so fanatical about this that the desire outweighs all of the future desires in L2 (or at least all of the future desires in L2 that get satisfied). Even if the father's knowledge of all of these desires caused corresponding present desires in him – one for each and every future desire – this still wouldn't be enough to tilt his motivation away from L1. According to Simple Desire Satisfactionism, then, the value of satisfying this desire is likewise through the roof. According to Simple Desire Satisfactionism, L1 is actually better for the father than L2, despite L2's being so much longer. Perhaps the objector will regard this as implausible; if the dedicated dad is sacrificial, so must be the fanatical father.

But this judgment is far from clear. Consider the case abstractly: L1 and L2 are bundles of goods; L2 contains a far greater number of goods than L1; but L1 contains a good of tremendous value, so tremendous that its value outweighs the sum of the values of all the goods in L2.<sup>13</sup> So it's better to get L1 than it is to get L2. So a choice of L1 over L2 is no act of self-sacrifice.

---

<sup>13</sup> I'm fudging a little here for simplicity's sake. L1 and L2 are entire lives; it is no part of Fanatical Father that the tremendously strong desire in L1 outweighs all *past*, *present*, and future desires in L2. Only future desires are at issue. To remedy this, let

As noted above, Fanatical Father is actually a special case – and an especially exotic one – of the third and final objection I will consider. We might do better to consider the latter.

This third objection admits that Simple Desire Satisfaction does allow for *some* cases of self-sacrifice, but not enough. There are, the objection contends, cases of self-sacrifice that the theory does not recognize. I will flesh out the objection in a moment, but notice that, even if the objection were sound, the main point of this paper would still stand: the most simple, unrestricted form of the desire theory – so long as it is atomistic – would be compatible with self-sacrifice, even if it doesn't recognize enough of it. Nevertheless, I think the objection can be answered.

The case above of the suicidal father is, in a sense, “diachronic”: self-sacrifice occurs because the father chooses today to forgo goods he would receive in the future. But consider a more “synchronic” case, a case in which someone chooses today to forgo good she would receive today. Suppose Emma is deliberating over how to spend her Friday night. She can either go to the disco with her friends, or she can volunteer at the local soup kitchen. Emma considers the options and decides it would be best to spend her Friday night helping the needy. She feels it would be the right thing to do. So this is what she most prefers to do and she does it.

The objection I have in mind claims that Emma's act should count as an act of self-sacrifice. After all, she sacrifices a fun night out with her friends in order to help the needy. The claim is not that Emma is forgoing future goods down the road. We can

---

F1 be the future segment of L1 and F2 the future segment of L2. Substitute ‘F1’ for ‘L1’ and ‘F2’ for ‘L2’ in the above.

assume that whatever Emma chooses today will have no relevant effect on what happens in the future. So the idea is that Emma is sacrificing something good for herself today (viz., fun) for duty's sake. In the case imagined, the only way to distinguish Emma's two available lives is this Friday night. But the possible Friday night in which she volunteers at the soup kitchen contains more desire satisfaction for her than does the Friday night in which she goes to the disco. For, we can suppose, at each moment at which she is at the soup kitchen, she desires to be there more strongly than she desires to be at the disco. Indeed, we need this supposition if we are to force the Simple Desire Satisfactionist into denying self-sacrifice in this case.

So the argument against Simple Desire Satisfactionism is this. If the theory is true, then the Friday night in which Emma volunteers at the soup kitchen is better for Emma than the Friday night in which she goes to the disco with her friends (for it better satisfies her desires). Since the two corresponding lives differ in this way only, Emma's volunteering at the soup kitchen is in her best interest. Therefore, doing so is not an act of self-sacrifice. But, claims the argument, this *is* an act of self-sacrifice. So Simple Desire Satisfactionism is false.

But I don't think it is implausible for the Simple Desire Satisfactionist to stand his ground here and deny that Emma's act is a net self-sacrifice. Emma has a stronger desire to help the hungry than to go out with her friends. As stipulated, the rest of her life will not be affected in any way by these decisions. I don't see why we must say that Emma is sacrificing herself. It is question-begging to appeal to the fact that going to the disco would be more fun. That just seems to presuppose a hedonist conception of welfare. It is not a hidden or unintended consequence of the desire theory that things

can go well for an agent without things being fun for the agent. Emma spends her Friday night doing what is most important to her, what she most wants to do. She sacrifices no future goods to do this. I say she did what was most in her interest on that Friday night.<sup>14</sup>

### 3.6 Closing Remarks about Simple Desire Satisfactionism

Even if no self-sacrifice argument undermines Simple Desire Satisfactionism, the theory may stand in need of some refinement. But its simple form allows for straightforward revision, and since the theory can remain atomistic and summative, such revision will not threaten the theory's capacity to handle self-sacrifice. Maybe, for example, we should prefer a Parfitian hybrid theory and reduce or eliminate the value of a desire satisfaction if the object desired is not found on the objective list (Parfit, 1984, pp. 501-502). Or maybe we should regard the satisfaction of "global desires" to be more valuable than local ones.<sup>15</sup> This would enable the theory to co-opt some of the appeal of holistic theories. Maybe we should count the satisfaction only of desires about things that happen while the subject is still alive. Or maybe we should go further and count a desire satisfaction only if the subject is aware that he has gotten what he

---

<sup>14</sup> I thank an anonymous referee for the *Canadian Journal of Philosophy* for raising an objection of this sort.

<sup>15</sup> See Parfit (1984, pp. 496-498). Parfit puts global theories in opposition with summative theories. But as I use the terms, nothing prevents a global theory (one in which the value of the satisfaction of a global desire is enhanced) from being a summative theory (one that determines the value of a whole life by summing over the values of the atoms).



wanted.<sup>16</sup> We could even follow conventional wisdom and count rational instead of actual desire satisfactions without sacrificing the ability to handle self-sacrifice. Can the atomist framework accommodate each of these intuitions? Yes it can. Should it accommodate them? That's a question for another time.

---

<sup>16</sup> This may be a better solution to Parfit's (1984, p. 494) "stranger on the train" case than the one Parfit himself suggests.

## CHAPTER 4

### SUBJECTIVE DESIRE SATISFACTIONISM

In this paper, I offer a theory of welfare: an answer to the ancient question, What makes a person's life go well for him or her? According to what is probably the reigning view (or at least the reigning view is that it's the reigning view), welfare has to do ultimately with *desire*. I agree, but I don't think just what welfare has to do with desire has been properly recognized. The best life for you is not the one you want most; nor is it the one you would want most if you were somehow improved (e.g., by becoming fully informed about and/or fully rational and/or fully imaginatively acquainted with each of your possibilities); nor is it the one some ideal observer, or God, would want for you. The best life for you is not the one that best satisfies the desires you would have if you were to lead that life (though, in my view, we're getting warmer). Nor is the best life the one that best satisfies just your global desires, or just the desires that are about your own life. All of these theories are open to decisive objections. In my view, the best life for you is the one that best "subjectively satisfies" the desires you would have if you were to lead that life. It is not the life in which you are getting what you want, but in which you *think* you are getting what you want. This is *Subjective Desire Satisfactionism*.

Subjective Desire Satisfactionism is a theory of welfare, a theory of intrinsic value for a subject, a theory about what makes a state of affairs intrinsically good (or

bad) for a subject. This is different from a theory of intrinsic value *simpliciter*, which tells us what makes a state of affairs intrinsically good – not for any subject, but just plain good. The theory of welfare tells us what makes a life a good life for the one who lives it. But it need not be a theory about the other ways in which a life can be evaluated: it is not a theory of the *moral value* of a life; it's not a theory of what makes a life *virtuous* or *vicious*; it is not a theory of the level of *dignity*, or *excellence*, or *achievement* in a life; it is not a theory of the *beneficialness* of a life (i.e., how well off the person makes those around him); it is not theory of the *contributory value* of a life (i.e., what direct contribution the existence of the life makes to the value of the world); and it is not a theory of the *aesthetic* or *narrative value* of a life (roughly, how good the story would be if the life were made into a story). It is important to keep these distinctions in mind; in my view, some arguments against desire theories of welfare rest on a conflation of these different scales of evaluation.

I begin by surveying the main kinds of desire-based theories of welfare and presenting the arguments that some think show them to be unsatisfactory. Then I present Subjective Desire Satisfactionism and explain how it handles all of the these, and other, objections.

#### **4.1 Holistic Desire-Based Theories of Welfare**

I find it natural and useful to divide desire-based theories of welfare into two fundamental kinds: *holistic* theories and *atomistic* theories. Holistic theories are (or at least tend to be) archetypes of *subjectivism about welfare*: according to them, how good your life is for you is to be explained in terms of your (or some ideal observer's)

attitude towards that life as a whole. According to holism, the amount of welfare in a life is not derived simply by adding up the amounts of welfare in some specified parts of the life. Atomism about welfare is the opposite: according to atomism, there are minimal bits of welfare; there are good bits and bad bits; and they accrue over time in a life to constitute how well the person's life went for him. Atomistic theories are (or at least tend to be) *summative*: to get the value of a life, simply add up the values of all the minimal bits of welfare. Atomistic theories are *objectivist*, at least in one sense: they don't care how you (or anyone) feels about the life as a whole; their recipe for determining the welfare value of a life might rate your life very high even though you think the life as a whole is not very good (and vice versa: it can deem a life a poor one even though its liver thinks it pretty good).

#### **4.1.1 Naive Preferentism and the Arguments from Ignorance, Weakness of Will, and Unimaginativeness**

I begin with holism/subjectivism.<sup>1</sup> It's natural to state such theories as forms of *preferentism*, i.e., as theories according to which welfare has ultimately to do with facts about what subjects prefer. (However, since I assume that preference is definable in terms of desire, preferentist theories do qualify as desire-based theories of welfare.) The most anti-paternalistic among us might be tempted to think that what makes one life a person could lead better than another is that the person herself prefers it to the other. *You* get to say which life is best for you. Call this theory 'Naive Preferentism'.

---

<sup>1</sup> The holism-atomism distinction and the subjectivism-objectivism distinction probably cut across each other. Nevertheless, I'm here going to treat holistic and subjectivist theories together. I'm confident no winning theory will fall through the cracks.

If L1 and L2 are two possible lives that a subject S might lead, we can state the theory as follows:

**Naive Preferentism:** Life L1 is better for a subject S than life L2 iff S prefers L1 to L2.

This theory deserves its name mainly because a person's actual preferences concerning her possibilities might be based on ignorance, unimaginativeness, or irrationality, which could lead her to prefer a life even she would recognize as worse. Ignorant Ignacio might prefer the life in which he drinks from the river now to the life in which he waits to get back to camp to quench his thirst.<sup>2</sup> Unknown to Ignacio, the river is poisoned and will make him sick for months. Naive Preferentism implies that Ignacio is better off in the sick life and that we therefore don't benefit him by warning him about the river. But of course we do.

Weak-willed Willie might prefer the life in which he avoids the dentist to the life in which his suffering is relieved – but only because he is dentophobic: he has an irrational fear of dentists and would go through just about anything to avoid them.<sup>3</sup> To insist that whatever Willie prefers most must be best is liberalism gone mad.

Unimaginative Maggie might know all about one of her possibilities: the life of a paparazzied pop star. But she might fail to appreciate what she knows; she might not imagine well enough just what it will be like to be pursued incessantly by the paparazzi. Her lack of imagination and appreciation, rather than her lack of knowledge or will, might lead her to prefer what is worse for her.

---

<sup>2</sup> This case is like one from Carson (2000, pp. 72-73). A relevantly similar case appears in Schwartz (1982, pp. 195-196).

<sup>3</sup> This case is like one from Schwartz (1982, p. 196).



#### 4.1.2 Ideal Preferentism

Cases such as these have led many philosophers to reject a view like Naive Preferentism but without abandoning the general holistic/subjectivist approach. They modify the theory by *idealizing preferences*. The most standard kind of idealized preference is the *fully informed* preference. But full-information might not solve the case of unimaginative Maggie, and so some theories also require *vividly keeping the information in mind* or *full imaginative acquaintance* with the content of the information.<sup>4</sup> The preference is ideal only if the subject really imagines what the life would be like; and in order to do this, the ideal subject might even have to have experienced the same sorts of experiences found in the life. Let's call a person 'fully and vividly informed' about some possibility if she is fully informed about it, keeps the information about it vividly in mind, and has full imaginative acquaintance with respect to it. We have, then,

**Ideal Preferentism:** Life L1 is better for a subject S than life L2 iff S would prefer L1 to L2 if S were fully and vividly informed about L1 and L2.

Sidgwick (1907, pp. 111-112) suggests a theory quite like Ideal Preferentism in his famous remark that

a man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realised in imagination at the present point of time.

It is worth pointing out that in moving to Ideal Preferentism, some of what has attracted many to the desire approach to welfare has been lost. Ideal Preferentism, for example,

---

<sup>4</sup> See Sidgwick (1907, pp. 111-112) and Brandt (1972, p. 682).

is far less liberal than Naive Preferentism. The dictates of one's ideal self might seem to one as oppressive as those of any pater. Furthermore, Ideal Preferentism is no longer empirically respectable – what reliable experiment can test such exotic counterfactuals?<sup>5</sup>

Ideal Preferentism sounds plausible enough. When a person finds himself at a major crossroads in his life – perhaps in deliberation about whether to marry, or which career path to choose – it is not uncommon, and usually recommended, to engage in a procedure that is suggested by Ideal Preferentism. The person is deciding between different ways his life could go. It doesn't seem unreasonable that, to make the best decision, he gather as much information as he can about each possible life, that he use this information to vividly imagine what it would be like to live each of the lives, and that, after doing so, he decide which life he most prefers to lead. This common practice seems to be an unconscious endorsement of something like Ideal Preferentism.

But I don't know how Ideal Preferentism is supposed to handle weakness of the will. We can assume that weak-willed Willie is fully and vividly informed about each of his alternatives – that is, that his preferences are ideal. We can even suppose that he admits that he would be better off in the life in which he goes to the dentist. But, nevertheless, he can't bring himself to go – he voluntarily and knowingly chooses not to go. Therefore, he prefers not to go.<sup>6</sup> Therefore, his ideal preference is for what is

---

<sup>5</sup> See Sumner (1996. pp. 119-120) and Adams (1999, pp. 86-87).

<sup>6</sup> We are therefore working with a broad notion of desire and preference, one according to which, necessarily, if a person voluntarily and knowingly chooses to bring about state of affairs p over state of affairs q, then *ipso facto* he prefers p to q.

worse for him. Therefore, even Ideal Preferentism goes wrong with respect to the case of weak-willed Willie. This again is the *Argument from Weakness of the Will*.

#### 4.1.3 The Argument from Self-Sacrifice

There are more troubling problems with Ideal Preferentism. The *Argument from Self-Sacrifice*, due to Overvold (1980), shows that such theories are incompatible with the possibility of self-sacrifice. That argument is complex; it is simpler to show how the theory goes wrong with respect to a particular case of self-sacrifice. Overvold (1980, p. 108) invites us to

Suppose a man wants more than anything else that his four sons attend a very expensive private college. He is not a rich man. The closest thing to a tangible asset he possesses is a huge life insurance policy. After carefully considering his options, he resolves to kill himself, making it look like an accident. He does so, and four years later his eldest son begins college. Eventually all four sons complete their education and enjoy very happy and rewarding lives.

Let's suppose that had the father decided against suicide, he would have eventually "gotten over" his inability to send his sons to college. He would have gone on to lead a long and overall reasonably satisfying life, a life well worth living by all accounts. Nevertheless, the fully informed and imaginative father prefers and chooses the life in which he commits suicide. It is obvious that the correct way to describe this is as an act of self-sacrifice. The man sacrifices a long and desirable life for himself and takes instead a life cut short, all for the sake of his children's welfare. If we suppose that the suicide life is the one the father *most* prefers from among *all* his alternatives, Ideal Preferentism implies that the suicide life is in the father's best interests, and therefore that the act that brings about this life maximizes his total welfare. Since no act that

maximizes the agent's total welfare can be an act of self-sacrifice, Ideal Preferentism implies that the man's act of suicide is not self-sacrificial. But of course it is; if we know anything about self-sacrifice, we know that this father has done it.

There are many other contexts in which a person might, even with full information and full imaginative acquaintance, sacrifice her welfare. She might prefer what is worse for herself for the sake of her interests in the truth, or for the sake of duty, or in service to a noble cause.<sup>7</sup> These considerations give rise to alternative versions of the Argument from Self-Sacrifice.

#### **4.1.4 Egocentric Ideal Preferentism and Its Difficulties**

It is tempting to reply that, while perhaps in these cases the subject prefers the worse life "all things considered," it is not the one he prefers "for himself." What the father prefers is quite an extensive state of affairs, one including how well his kids are doing; his own life is just one small part of the scenario. Just thinking of himself and his own life, the man would (we can suppose) prefer not to commit suicide. This reply concedes the success of the Argument from Self-Sacrifice and offers a restricted theory in its place. Sidgwick may have had the kind of restriction in mind when he instructs us to "consider only what a man desires ... for himself – not benevolently for others" (1907, p. 109). And Overvold's suggested revision, which seeks to count only desires

---

<sup>7</sup> See Adams (1999, pp. 87-89).

that “directly concern the agent,” is along exactly these lines. So is Parfit’s suggestion that we count only desires “about our own lives” (1984, p. 494).<sup>8</sup>

Only Overvold provides a clear way to understand what it is for a desire directly to concern the agent. He states his theory as follows (1980, p. 117-118, 10n):

an act is in the agent’s self-interest if and only if it is the act which the agent would most want to perform if he were (a) fully aware of all the features and outcomes of the alternative acts open to him at the time and (b) his choice were motivated only by his rational desires and aversions for features and outcomes of the act which are such that the proposition asserting that the agent exists at *t* is a logically necessary condition of the proposition asserting that the feature or outcome obtains at *t*.<sup>9</sup>

The idea is that, to find out which of a person’s possible lives is best for him, we find out which of them he would prefer most if (a) he were fully and vividly informed about each possible life, and (b) he had only desires whose content entails his own existence.

Call desires of this sort *egocentric desires*.<sup>10</sup> We can state the theory as follows:

**Egocentric Ideal Preferentism:** Life L1 is better for a subject S than life L2 iff S would prefer L1 to L2 if S were fully and vividly informed about L1 and L2 and had only egocentric desires.

It is important to recognize that Egocentric Ideal Preferentism is not circular, though it may appear at first glance to be. For it is tempting to understand an egocentric desire to

---

<sup>8</sup> Though Parfit is moved to make this restriction not from cases of self-sacrifice but from the case of the stranger on the train. More on this case later.

<sup>9</sup> The theory is presented as a theory of self-interest rather than as a theory of betterness of life. But the two kinds of theory are intertranslatable, as follows: an act is in an agent’s self-interest iff the life he would lead if he were to perform that act is at least as good as the life he would lead were he to perform any alternative to that act.

<sup>10</sup> Overvold seems to allow the person to *have* any desires but then restricts which desires can *motivate* him. Since the notion of a non-motivating desire sounds dubious, I require instead that the person have only egocentric desires.



be a self-interested desire, a desire for one's own welfare. But this is not the concept; 'egocentric desire' is analyzed without appeal to the concept of welfare, self-interest, or goodness for a subject.

Egocentric Rational Preferentism does allow for self-sacrifice. Return again to the case of the suicidal father. If the father had only egocentric desires, he would have no desire for his kids to go to college (his kids's going to college does not entail his own existence). We can stipulate that if in fact the father were to rid himself of all his non-egocentric desires, his desire for the longer life would be stronger than his desire for the suicide life. Egocentric Rational Preferentism would then imply that the former is better, which it is.

But Overvold's theory faces new difficulties. One strikes at the very heart of the refinement of its predecessor. It turns out upon examination to be very implausible to suppose that no desire that does not entail my existence can be relevant to my welfare. Suppose I have strong desires for philosophical fame: I want my papers to be widely read and highly regarded. These desires are not egocentric – their satisfaction is compatible with my non-existence. Suppose I make this my life's work, succeed at it, relish in it, and as a result become thoroughly satisfied with my life. Egocentric Ideal Preferentism implies that we should ignore all of this when assessing my welfare (so long as the relishing and the satisfaction occurs only in virtue of my desiring the fame, which it certainly might). It implies that I could have been better off by attending to things I care far less about, so long as these things entail my own existence. Call this the *Argument from Important But Non-Egocentric Desires*.

Notice that the case under consideration is not the more familiar (but in my view less compelling) case of the achievement of posthumous fame. In the case at issue, I desire and get and recognize and enjoy the fame while I am alive. Egocentric Ideal Preferentism still tells us to rate my life from the perspective of someone who didn't have the desire.

What the case shows, I think, is that Overvold's analysis of what makes a desire "directly concern the agent" is faulty. I think, intuitively, my desire that I be famous *does* "directly concern me," even though it doesn't entail my own existence. Maybe when we say that some desire directly concerns some agent we just mean that its satisfaction is relevant to his welfare. If so, then this preferentist program is a dead end.

Another difficulty for Egocentric Ideal Preferentism is that although it allows for some self-sacrifice, it doesn't allow for enough: there are obvious acts of self-sacrifice that the theory does not recognize as such. I am thinking of cases in which a person has a desire that we would classify as a not *purely* self-interested desire, but whose content entails the desirer's existence. Such cases show that egocentric desires don't perfectly track self-interested desires. When these concepts come apart, we can have self-sacrifice that Egocentric Ideal Preferentism doesn't recognize. For example, I might desire not merely that the hungry be fed but that *I* feed the hungry. If I knowingly suffer personal losses towards this end, I am sacrificing myself, contrary to the opinion of Egocentric Ideal Preferentism.<sup>11</sup> Call this the *Argument from Egocentric Self-Sacrifice*.

---

<sup>11</sup> Cf. Adams (1999, p. 88).

Although I think the considerations already discussed are enough to do away with Ideal Preferentism and Egocentric Ideal Preferentism, I want to finish this section on holism by raising several other arguments. It will be useful to have these arguments on the table when later evaluating Subjective Desire Satisfactionism. I'll present these as arguments against Ideal Preferentism though each applies equally well to Egocentric Ideal Preferentism.

#### 4.1.5 The Argument from Self-Loathing

The *Argument from Self-Loathing* is based on the idea that people who hate themselves, or have sufficient amounts of guilt, can prefer (even when fully and vividly informed) that they be worse off. A person might believe that one of his options would be the worst for him and, for that very reason, choose it. Richard Kraut (1994, p. 165) has us

... suppose a man has committed a serious crime at an earlier point in his life, and although he now regrets having done so, he realizes that no one will believe him if he confesses. So he decides to inflict a punishment upon himself for a period of several years. He abandons his current line of work, which he loves, and takes a job that he considers boring, arduous, and insignificant. He does not regard this as a way of serving others, because he realizes that what he will be doing is useless. His aim is simply to balance the evil he has done to others with a comparable evil for himself.<sup>12</sup>

We can suppose that this person is fully informed about the life he prefers and is able fully to imagine what it would be like. If either Ideal Preferentism or Egocentric Ideal Preferentism were true, then no matter what horrible things this man filled his life with,

---

<sup>12</sup> What I call the Argument from Self-Loathing is discussed also in Adams (1999, pp. 89-91) and elsewhere.

so long as he foresaw them (and “forefelt” them, to use Sidgwick’s phrase) and continued to prefer them, his life could rank high in terms of welfare. The theory implies that, so long as this guy is fully informed and imaginative enough, he can’t succeed in punishing himself. What an odd implication.

#### **4.1.6 The Argument from Changing Preferences**

As it stands, Ideal Preferentism implies that judgments of betterness of life are not absolute, and that instead they are relative to this or that time. According to the theory, there is ultimately no absolute fact of the matter about whether some life is better than another. The best we can do is say is that this life is better than that life *at some particular time*. At some other time, that life might be better than this life. Does this even make sense?

Ideal Preferentism says that a possible life is better for someone than another just in case he ideally prefers the one over the other. But people have preferences at times. I might prefer one life over another at one time and then reverse the preference at a later time. Now, surely one’s informed preferences will in typical cases be more stable than one’s actual preferences, since one common cause of preference change is knowledge acquisition. But that’s not the only factor, and it takes only the smallest amount of variation over time to commit Ideal Preferentism to temporal relativism.

It seems possible, for example, that one’s rational preferences could change with changes in one’s mood. To locate an actual person’s fully and vividly informed counterpart, we change only the knowledge the actual person has and his level of appreciation of that knowledge. We don’t change his character, personality, or mood.

Consider again the suicide case. Suppose the father – who, recall, has informed preferences concerning the decision in question – has one week to deliberate over his decision. Suppose on Monday he feels magnanimous and prefers the suicide life to the longer life. By Tuesday he finds himself filled with spite and can muster little interest in his children's welfare; he prefers the longer life. On Wednesday, he wakes up in a loving mood and returns to his earlier preference for the suicide life. On Thursday, he is overcome with fear at the thought of his own demise and cannot bring himself to prefer suicide. Throughout all of these changes, of course, the man kept all the available information vividly before his mind and remained imaginatively acquainted with each life. It's just that, given the vagaries of his mood swings, the information and the imagining affect him differently on different days.

With respect to this case, Ideal Preferentism implies both that the suicide life is better than the longer life and that the longer life is better than the suicide life. Unless these verdicts are relativized to times, Ideal Preferentism therefore implies a contradiction. The current formulation appears to presuppose a falsehood – that ideal preferences cannot change over time – and the most natural correction is the following:

**Temporally Relativized Ideal Preferentism:** Life L1 is better *at time t* for a subject S than life L2 iff S would *at t* prefer L1 to L2 if S were fully and vividly informed about L1 and L2.

Temporally Relativized Ideal Preferentism has the odd result that on Monday, the suicide life is better for our suicide contemplator, but that on Tuesday the longer life is better. And then on Wednesday the suicide life returns again as the life with more welfare, but then on Thursday the longer life becomes most in his interest. This is difficult to swallow. The subject's own recognition of the vagaries of his preferences



might cause him to think, “Some days I prefer this life, other days I prefer that life, but I wonder which life really is best for me? Which one *should* I choose? *That’s* the one I shall choose.” Given the theory, this question has no satisfying answer.

Temporally Relativized Ideal Preferentism appears to have another unacceptable consequence. It appears to imply that, after a person dies – or at least long after he dies – there may be no fact of the matter about the relative value of any of the lives he might have led. Consider the life Socrates led. Now imagine a life he could have led if he had decided to become a carpenter instead of a philosopher. Fill in the details any way you like. Now we ask which of these lives is (now<sup>13</sup>) better for Socrates. Ideal Preferentism says that the answer depends upon which of these two lives Socrates (now) would have preferred, had he been fully and vividly informed (now) about both. Suppose Socrates, like the moody father discussed above, constantly changed his ideal preferences. Now consider a counterfactual of the form “If Socrates were alive today, and were fully and vividly informed, he would prefer ... .” On the supposition that Socrates was sufficiently fickle, I don’t see how any counterfactual of that form could be determinately true or false. Which Socrates do we consult? There would seem to be no non-arbitrary answer. Ideal Preferentism therefore implies that there is no fact of the matter about whether the possible life you imagined for Socrates would have been

---

<sup>13</sup> Whatever that means. But Temporally Relativized Rational Preferentism entails that the temporal index here is significant.

better or worse for him. Likewise for many such lives Socrates might have led. Call these last two arguments *Arguments from Unacceptable Indeterminacy*.<sup>14,15</sup>

#### 4.1.7 The Argument from Unwanted Ideal Lives

James Griffin raises the following argument against the “informed-desire account” of welfare. Though his target appears to be an atomistic rather than a holistic theory, the argument appears to make equal trouble for Ideal Preferentism. Griffin writes (1986, p. 11):

It is doubtless true that if I fully appreciated the nature of all possible objects of desire, I should change much of what I wanted. But if I do not go through that daunting improvement, yet the objects of my potentially perfected desires are given to me, I might well not be glad to have them. That is true, for instance, of acquired tastes; you would do me no favour by giving me caviar now, unless it is part of some well-conceived training for my palate.

Suppose we were to grant this much: that if my preferences were to be idealized in the manner of Ideal Preferentism, then the life that is ranked highest by those preferences

---

<sup>14</sup> Unacceptable because surely *some* amount of indeterminacy in welfare judgments is to be expected. For example, there surely are borderline cases in which there is no fact of the matter about whether a person desires (or would desire) some state of affairs. Thus *any* desire theory would seem to be subject to some indeterminacy. The claim here is just that the indeterminacy of Temporally Relativized Ideal Preferentism rises to an unacceptable level.

<sup>15</sup> Different forms of preferentism – such as those that consult the preferences not of one’s idealized self but of God, or some other ideal observer – may avoid some of these arguments from changing preferences. On the other hand, maybe not – is it really metaphysically impossible for an ideal observer to change her mind? The view of Bricker (1980), which looks at all of our changing preferences throughout our whole life and constructs a “timeless perspective” from which to assess our available lives, may also avoid some of these arguments (that is, if we understand it as a view about welfare as well as what it is explicitly a view about: prudence.)

would be best *for my idealized self*. We still wouldn't have granted the truth of Ideal Preferentism, because that is a theory about what is good not just for my idealized self but for me as I actually am. What if the life most preferred by my ideal self would be unbearable for my actual self?

Proponents of theories like Ideal Preferentism often assume that one's fully and vividly informed preferences are one's "true preferences." It is assumed that these preferences reflect what one "really wants." The point of these vague ideas may be to suggest that if I were presented with the objects of my ideal preferences, I would necessarily accept and embrace them: "Yes, *this* is what I've always been after, though I didn't always know it." For all I know, this is what would happen in typical cases. But it certainly need not happen. There can be no necessary connection between a person's getting the objects of her ideal preferences and her liking it, for what could account for such a connection? There is nothing incoherent about a case in which a person hates every minute of her life even while that life is ranked highest by her ideal preferences. Since it is a truism that no life that one hates every minute of is a good one, the *Argument from Unwanted Ideal Lives*, as I will call it, poses a problem for Ideal Preferentism.

Peter Railton has suggested a way to revise ideal desire theories so as to avoid this argument (1986, p. 16).<sup>16</sup> Don't look to which life your ideal self prefers; rather, look to which life your ideal self would prefer *for you yourself as you actually are*. The theory, which I'll call 'Reflexive Ideal Preferentism', can be stated most easily if we

---

<sup>16</sup> Or at least he has suggested a theory that itself suggests a way to revise Ideal Preferentism.

give a name to a concept we have already been using, the concept of an *ideal counterpart*.<sup>17</sup> A person's ideal counterpart is the person one would be if one were fully and vividly informed about one's possibilities. According to Ideal Preferentism, the better life for you is the one your ideal counterpart prefers. But according to

**Reflexive Ideal Preferentism:** Life L1 is better for a subject S than life L2 iff S's ideal counterpart prefers that S lead L1 rather than L2.

But what ensures that your ideal counterpart's preferences for you will be good for you? Who knows what your ideal counterpart would prefer for you? Maybe she will have your welfare in mind, but maybe she won't. Maybe she will find your ignorance and inexperience pathetic and feel only disdain for you, and therefore wish you ill. There is nothing inconceivable about this, and so Reflexive Ideal Preferentism does not reveal the nature of welfare. Call this the *Argument from Malicious Ideal Counterparts*.

This complaint immediately suggests an emendation of the theory, one that makes use of the concept of a *benevolent ideal counterpart*. A person is your benevolent ideal counterpart only if his preferences are ideal *and he desires that you be well off*. The circularity of such an approach should be obvious.<sup>18</sup>

---

<sup>17</sup> By using the term 'counterpart' I do not mean to presuppose any particular theory of de re modality.

<sup>18</sup> The theory defended by Thomas Carson (2000, pp. 239-254) puts God in the role of the benevolent ideal counterpart. But I don't think his attempt to define a *loving* God without appeal to the concept of welfare is successful.

#### 4.1.8 The Argument from Unworthy Desires

There is something attractive about the idea that some things, just by their nature, simply shouldn't be preferred. The idea is that some states of affairs are intrinsically unworthy of being desired. If you believe in intrinsic value, and in particular you believe that some states of affairs are intrinsically bad – intrinsically bad *simpliciter* and not just intrinsically bad *for* someone – then I think it is not much of a step to believe that some states of affairs are intrinsically unworthy of being desired. What *is* a substantial step, in my view, is the step from the claim that some state of affairs is unworthy of being desired to the claim that getting that state of affairs is bad for the person who desires it.

What are the allegedly offensive desires? There are *malicious desires*, which range from actively desiring to bring harm to others, as apparently exemplified by, for instance, the Washington D.C.-area sniper, to passively desiring to see others suffer, as illustrated by those who, say, secretly hope their friends fail. There are *degrading desires*, which Moore's (1993, pp. 146-147) case of the "perpetual indulgence in bestiality" exemplifies, provided we modify it to apply to desire theories of welfare rather than to hedonistic theories of value. There may be *tasteless desires*, which demonstrate a lack of aesthetic taste, as when a person prefers Muzak to Mozart, as well as a lack good taste generally, as when we prefer bad food or bad TV. Finally, there may be *pointless desires*. John Rawls's case (1971, pp. 432-433) of the talented mathematician who most prefers to spend his life counting blades of grass illustrates pointless desire. What he prefers doesn't strike us as particularly immoral or degrading or tasteless, but it does strike many as a decidedly unworthy of pursuit.



To get the Argument from Unworthy Desires to apply to holistic theories like Ideal Preferentism, we can imagine a fully and vividly informed person surveying his possibilities. One possibility is a life we would consider more ordinary – the life, say, of a teacher who enjoys meaningful friendships, the love of a devoted spouse, a few engrossing and worthwhile hobbies, and reasonable career accomplishments – a fine life by all accounts. Let the alternative possibility be the life of a sniper on the run, successfully terrorizing the community, endlessly relishing in the suffering of others, while continually eluding the authorities. Or let it be a perpetual indulgence in bestiality. Or let it be a life filled with otherwise unworthy pursuits. Full information about and full imaginative acquaintance with each of these lives certainly won't guarantee an aversion to them – indeed, in some cases it might be exactly what is responsible for a preference for them. Ideal Preferentism therefore implies that, in some cases, the bestiality life or the sniper life (or some other unworthy life) is a better life than the teacher life.

My own view is that it is not so implausible to “bite the bullet” in the face of the Argument from Unworthy Desires. Concede that things are going well for the sniper or the indulger. The bite can be taken out of this admission if we take care to distinguish the different scales on which lives can be ranked, and notice how low these lives rank on these other scales. Perhaps our disapproval of them can be fully accounted for by the realization that they rank low on some of the other scales on which lives can be measured (such as the moral goodness scale, the virtue scale, or the dignity scale), even if they rank high on the welfare scale.

## 4.2 Atomistic Desire-Based Theories of Welfare

I said that, according to atomism, there are minimal bits of welfare, good and bad bits, and that they accumulate over time to make up one's overall well-being. We better add that the bits have to be smaller than whole lives; otherwise holism is atomism plus the thesis that lives are the atoms. It is also worth pointing out here that nothing bars a holist about welfare from deeming smaller parts of lives as having some welfare value. Their defining claim is just that the value of a life is not a function of the values of these, or any, of its smaller parts; lives are "organic unities."<sup>19</sup>

### 4.2.1 Simple Desire Satisfactionism

Whereas holistic desire theories of welfare maintain that the welfare value of a life is determined by desires *about* that life as a whole, from a "bird's eye view" as it were, atomistic theories (at least the ones on which I will focus) maintain that the value is determined by the desires *within* that life. The simplest such theory postulates instances of desire satisfaction to be the positive atoms of welfare. These are states of affairs in which a subject desires that something be the case, and in fact it is the case. Desire frustrations, cases in which a subject desires something that is in fact not the case, are the bad atoms.

The theory should also say *how good* a given atom is. A simple and common idea is to make the value of a desire satisfaction or frustration a function of the *intensity* and the *duration* of the desire satisfied or frustrated. On this view, the longer a desire is

---

<sup>19</sup> This is not, by the way, to deny the supervenience of welfare upon all the other features of a life. Holism is not committed to that unbelievable thesis.

held, and the stronger it is as it is held, the better its satisfaction for its subject. If we don't mind assigning numbers to these quantities, we can say that the precise value of a desire satisfaction for its subject is the *product* of the duration of the desire and the average intensity of the desire over this duration (*average* since intensity can vary over time).

Finally, the theory should specify how to compute the welfare value of an entire life, or of a segment of a life, given the values of the atoms contained in that life or life segment. The simplest answer is: add them up. The theory, which I will call 'Simple Desire Satisfactionism', consists then of the following three clauses:

**Simple Desire Satisfactionism:**

- (i) Every desire satisfaction is intrinsically good for its subject. Every desire frustration is intrinsically bad for its subject.
- (ii) The intrinsic value for its subject of a desire satisfaction = the duration of the desire satisfied  $\times$  the average intensity of the desire satisfied. The intrinsic value for its subject of a desire frustration =  $-$  (the duration of the desire frustrated  $\times$  the average intensity of the desire frustrated).
- (iii) The intrinsic value of a life (or life segment) for the one who lives it = the sum of the intrinsic values of all the desire satisfactions and frustrations that occur in that life (or life segment).

In saying that every desire satisfaction is *intrinsically good* for its subject, clause (i) is identifying the atoms of welfare. On any atomistic theory, the atoms of welfare are *intrinsically* good for their subject; they themselves are the fundamental constituents of welfare. Something is *extrinsically* good for someone if it in some other way makes their life better than it would have been (e.g., by *causing* a state of affairs that is intrinsically good for him or her, or *preventing* a state of affairs that is intrinsically bad for him or her). Anything could be extrinsically good for someone (since anything can

cause anything), including the atoms of welfare themselves; but only the atoms can be intrinsically good for a person (in the most basic way).<sup>20</sup>

Simple Desire Satisfactionism is the simplest, most straightforward atomistic desire theory of welfare I can think of. It is completely *unrestricted*: it counts all desires, whether fully informed or ignorant, whether for the worthy or the unworthy, whether global or local, whether directly concerning oneself or purely about others. It is *actualist*: it counts all of the subject's *actual* desires, and none of his merely hypothetical ones. The theory is, in a sense, objectivist: how one feels about one's life does not settle the question of how good the life is. A person might have gotten just what he wanted on a day to day basis but then look back and be disappointed with his life. The theory can incorporate this disappointment, since it likely corresponds to a desire that his life had gone differently. But since it is easy for the satisfactions of his whole life vastly to outweigh this frustration, the theory implies that his judgment that his life was no good may be false. In another sense, the theory is still subjectivist: whether something is good for someone has fundamentally to do with the person's subjective attitudes. Though I do not endorse Simple Desire Satisfactionism, the theory I do favor also straddles the objectivist/subjectivist divide in this way. Perhaps this is what enables the theory, in my view, to steer a course between the pitfall of extreme subjectivism or (that whichever life the subject most prefers is automatically best) and the pitfall of extreme objectivism (that a person can (or can easily or commonly) get

---

<sup>20</sup> "In the most basic way" because we want to be able to say that composites containing intrinsically good atoms, including especially whole lives, can be intrinsically good (see Feldman (2000)). All of the above applies, of course, to intrinsic badness and the negative atoms of welfare as well.

what is intrinsically good for him and have no interest in it whatsoever). Whereas the former is liberalism gone mad, the latter paternalism gone mad.

The holistic preferentisms of the previous section were presented as theories of *betterness of life*. They give no absolute answer to the question of how good some life is; they tell us only which lives it is better than and worse than. This feature would sit well with those who believe that there is no such thing as the absolute welfare value of a life, and that instead we have only facts about the *comparative value* of two or more lives. Simple Desire Satisfactionism, by contrast, entails that there is such a thing as absolute welfare value. This is because it supposes there is such a thing as the absolute intensity to a desire (rather than merely *comparative* intensities between desires).

But Simple Desire Satisfactionism does of course entail a thesis of betterness of life: possible life L1 of subject S is better for S than possible life L2 of S iff L1 contains a greater balance of desire satisfaction over frustration than L2. The phrase ‘balance of desire satisfaction over frustration’ is meant to embody the fact that *number*, *intensity*, and *duration* of desire matter according to the Simple Desire Satisfactionism. The Simple Desire Satisfaction picture is this. There is a set of true propositions about the world. We have desires for many of these true ones as well as, unfortunately, desires for many of the false ones. Whenever we are desiring one of the true ones, we are accruing welfare. Whenever we are desiring one of the false ones, we are losing welfare. The total balance at the end of our life is how well things went for us. It is a stark, and perhaps simplistic, picture, but a useful place to begin. I have belabored its presentation because it is an ancestor of the theory I like.



So what is wrong with Simple Desire Satisfactionism? And how can it be improved?

#### 4.2.2 The Argument from Changing Desires

We saw that changing preferences raise the specter of relativism for holistic preferentism. Changing desires make trouble for atomistic desire satisfactionism in a different way. Consider

*Ellie's 50<sup>th</sup> Birthday Party.* Teenage Ellie, a rock 'n' roll fan, is imagining her 50<sup>th</sup> birthday party. She wants live rock 'n' roll at the party. She continues to desire for years and years that there be rock 'n' roll be at her 50<sup>th</sup> birthday party. But a month before the party, Ellie realizes she no longer enjoys rock 'n' roll. She now prefers easy listening, and she finds rock 'n' roll loud, childish, and annoying. She will continue to feel this way on her 50<sup>th</sup> birthday. She would have the time of her life at her party if she got easy listening, but would be miserable if rock 'n' roll were played.

Suppose we are planning Ellie's party, and we want to do whatever would be best for Ellie. Do we benefit Ellie more by giving her rock 'n' roll or easy listening on her 50<sup>th</sup> birthday?

It is obvious that we should give Ellie easy listening. This would be best for her. She'd be miserable at her own party if we forced rock 'n' roll on her; she no longer wants rock 'n' roll. If we asked her opinion, she would beg that easy listening be played. She'd plea, justifiably, "Forget about my past desires; give me what I want *now*, not what I used to want!"

But (so long as we fill in the numbers properly) Simple Desire Satisfactionism implies that we benefit Ellie more by force-feeding her the rock 'n' roll. Doing so will satisfy such a long-lasting desire, that the value of this satisfaction will far outweigh the value of satisfying her month-long desire for easy listening along with her desire not to

be miserable at her party. That Simple Desire Satisfactionism gives equal weight to merely past desires makes it, in some cases, as paternalistic as any objective list theory. Our past desires dictate what's good for us, whether we like it or not.

Perhaps there is a solution. Parfit says (1984, p. 151):

Some desires are implicitly conditional on their own persistence.  
If I now want to swim when the Moon later rises, I may want to do so only if, when the Moon rises, I still want to swim. If a desire is conditional on its own persistence, it can obviously be ignored once it is past.

This suggests an emendation of Simple Desire Satisfactionism: when a person has a desire about the future that is conditional upon its own persistence, its satisfaction is intrinsically good for the person when and only when the condition is satisfied – i.e., when the desire persists. Perhaps Ellie's original desire for rock 'n' roll was implicitly conditional upon its own persistence: as a teenager she wants rock 'n' roll at her 50<sup>th</sup> birthday party only if, on that day, she still wants rock 'n' roll. Since the condition isn't satisfied, satisfying the conditional desire doesn't benefit her.<sup>21</sup>

But I don't think this restriction will do the trick. For it could be that, as a teenager, Ellie finds easy listening so repulsive that her desire is not conditional upon its own persistence. Rock 'n' roll fanatics might say, "If ever I prefer easy listening, then shoot me." Ellie's case need not be so extreme: "Even if I prefer easy listening when I'm 50," she might say as a teenager, "I don't care – please still have rock 'n' roll at my party." Nevertheless, when the party is weeks away, and we're deciding on the music,

---

<sup>21</sup> It may be that the best framework in which to cash out the notion of a desire that is conditional on its own persistence is one that takes entire possible worlds rather than propositions to be the objects of desire. See Bricker (1980, pp. 388-89) Bricker employs a restriction along these lines in his desire theory of prudence.

it is still obvious that we benefit Ellie more by giving her easy listening. She'd plea, justifiably, "Forget about the fact that my past desire was not conditional upon its own persistence; I was foolish and dogmatic then; give me what I want *now*, not what I used to want!" This case convinces me that appealing to the notion of a desire's being conditional upon its own persistence doesn't solve the Argument from Changing Desires.

#### 4.2.3 An Argument from Anti-Paternalism

Although some degree of anti-paternalism in a theory of welfare is desirable, Naive Preferentism, we saw, was too anti-paternalistic. Simple Desire Satisfactionism might seem to err too far on the other end of the paternalism spectrum. The following case brings out the paternalism in Simple Desire Satisfactionism:

*A Brand New Life.* An eccentric billionaire with an experimental drug offers you a brand new life. You will be relocated to a new city – one that now doesn't appeal to you at all. You will be immersed in a new circle of friends – people with whom you now have no wish to associate. You will be given a new career – one you have absolutely no interest in. You will never be allowed to return to your current home, see your current friends or family again, or pursue your current career and other projects. But you will be given a drug – a complacency pill<sup>22</sup> – that will gradually cause you to want to be in your new city once you are there, and to want to associate with your new friends, and to want to be engaged in your new career. The drug will also cause your longings for your old life to diminish. As it happens, the life you will lead if you decline the offer has its ups and downs, and, although it is a fine life by any standard, has its share of desire frustration. The brand new life on offer, however, will fulfill far more of the desires you will have if you lead that life.<sup>23</sup>

---

<sup>22</sup> As in Bricker (1980, pp. 398-400).

<sup>23</sup> The drug changes you gradually rather than abruptly to ensure that you will survive the changes, in case some sort of psychological continuity view of personal identity is true.

Should you accept the billionaire's offer? Would it be in your best interest to accept the offer? Would you be foolish to decline it?

I suspect that very few of us who think our lives decent enough and feel attached enough to our friends, family, projects, and careers, would even contemplate such an offer. We recognize that we are turning down a life that would satisfy much more of the desires we would have if we were to lead that life. Nevertheless, we would be getting things we now care nothing about. The Brand New Life on offer is completely foreign to us. We're very happy with our current lives, thank you very much.

But Simple Desire Satisfaction implies that you'd be better off opting for the Brand New Life. It implies that you are "shooting yourself in the foot" by settling for the status quo. This seems objectionably paternalistic. The theory forces upon you a life you don't want. Don't *you* get any say in deciding which of these lives would be best for you?

Does Simple Desire Satisfactionism also imply that your choice is *irrational*? That you are a fool not to take your Brand New Life? That there's something wrong with you if you keep your present life? The theory on its own doesn't imply this. But it does imply this together with some auxiliary principles, such as those of the *Self-interest Theory*, discussed by Parfit (1984, pp. 4, 8, and *passim*):

(S1) For each person, there is one supremely rational ultimate aim: that his life go, for him, as well as possible,

or

(S2) What each of us has most reason to do is whatever would be best for himself,

or

(S3) It is irrational for anyone to do what he believes will be worse for himself.

If, say, (S3) is necessarily true, then Simple Desire Satisfactionism implies that it is irrational for you to decline the offer. If, however, it is clear to us that it is not irrational for you to do this, then Simple Desire Satisfactionism implies something that to us is clearly false.

Notice that the holistic forms of preferentism from the previous section avoid this result. They seem almost tailor-made to handle such an objection as they give weight only to your present preferences and not to any of the futures preferences within either of the lives on offer.

#### 4.2.4 The Argument from Remote Desires

The name I give this objection derives from the remoteness of the *objects* of desire. Sometimes we desire certain remote states of affairs to obtain, remote in time, in place, in importance. Derek Parfit (1984, p. 494) presents a version of this objection in the form of the case of The Stranger on the Train:

Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured. On the Unrestricted Desire-Fulfilment Theory, this event is good for me, and makes my life go better. This is not plausible. We should reject this theory.

Parfit's argument is pretty clearly meant to apply to atomistic theories. He talks of the event of the stranger being cured as being good for him and making his life go better.

Just to be clear, Simple Desire Satisfactionism does not imply that the stranger's being cured is intrinsically good for Parfit. The theory implies that a more complex state of



affairs is intrinsically good for Parfit: the conjunctive state of affairs of Parfit's intrinsically desiring that the stranger be cured and the stranger's being cured. In any case, Parfit is right – it seems quite implausible to suppose that this state of affairs, in part about something so remote to Parfit, makes his life go better.<sup>24</sup>

Parfit suggests a way to restrict the theory in question so as to avoid the objection: count only desires “about our own lives” (1984, p. 494). Parfit admits, however, that “when this theory appeals only to desires that are about our own lives, it may be unclear what this excludes” (1984, p. 494). We have already looked at one attempt to make this idea clear: Overvold's view that a desire is about one's life just in case it entails one's existence. But we also saw the implausibility of this restriction to the desire theory: it leaves out too much; desires other than those that entail my own existence are important too.

The Argument from Remote Desires is made also by James Griffin (1986, pp. 16-17):

The breadth of the account, which is its attraction, is also its greatest flaw. ... It allows my utility to be determined by things that I am not aware of (that seems right: if you cheat me out of an inheritance that I never expected, I might not know but still be worse off for it), but also by things that do not affect my life in any way at all. The trouble is that one's desires spread themselves so widely over the world that their objects extend far outside the bound of what, with any plausibility, one could take as touching one's well-being.

Griffin suggests a way to restrict the theory in question so as to avoid the objection: count a desire only if it is “one of [our] aims,” “one of [our] central ends” (1986, p.

---

<sup>24</sup> We of course are assuming, perhaps artificially, that Parfit's desire that the stranger be cured is intrinsic. But we could of course change the case to avoid this without its losing its force.

21).<sup>25</sup> If these notions can be cashed out in terms of the intensity of the desire, then the restricted theory can ignore, or at least discount the value of, the satisfaction of low-intensity desires. Griffin here is assuming that the force of the objection comes from the remoteness of the importance of the desire – they are remote to what we most care about.

But this doesn't seem right, as I think the following passage by Shelly Kagan (1998, p. 37) illustrates:

... according to the preference theory, I am made better off by the satisfaction of my various desires, regardless of the subject matter of the given desire. Suppose, then, that I am a large fan of prime numbers, and so I hope and desire that the total number of atoms in the universe is prime. Imagine, furthermore, that the total number of atoms in the universe is, in point of fact, prime. Since this desire is satisfied, the preference theory must say that I am better off for it ... . But this is absurd! The number of atoms in the universe has nothing at all to do with the quality of my life.

In Kagan's case the desire is, we can suppose, no passing whim. We can assume it is as intense and long-lasting as you like, and so does count as one of the subject's "central ends." If the objection had any force against the unrestricted theory, I think it has equal force against a theory restricted in the way Griffin suggests.<sup>26</sup>

---

<sup>25</sup> See also Scanlon (1998, p. 120-121).

<sup>26</sup> Scanlon (1999, ch. 3) endorses a "rational aim" theory over an informed desire theory, and, in doing so, distinguishes aims from desires. Above, I assume that aims are desires (though I do not assume that is all they are). Kagan's argument may not work against an "aim theory" that takes aims to be something other than desires.

#### **4.2.6 The Argument from Posthumous Harm**

Another complaint against an atomistic desire theory of welfare is that it allows for the possibility of posthumous harm. Some philosophers find it incredible that a person can be harmed after he is dead and gone. But Simple Desire Satisfactionism makes this possible. Suppose Billy really wants his Dodgers to win the World Series this fall, and suppose he dies even before spring training is over. The Dodgers make it to the series, but are thwarted again by those damn Yankees. We can say, “Poor Dodgers; too bad for them; they wanted it so bad, and it didn’t happen.” But if Simple Desire Satisfactionism is true, we can also legitimately say, “Poor Billy; too bad for him; he wanted it so bad, and it didn’t happen.” If standard desire theories are true, then not even when a man is dead is he beyond the reach of misfortune.

### **4.3 Subjective Desire Satisfactionism**

We have so far reviewed a litany of objections to desire-based approaches to welfare. I want to proceed by motivating and laying out my proposal. Then I’ll explain how this new theory handles the objections discussed above that need handling.

#### **4.3.1 The Experience Requirement**

The motivation for my proposal is simple. It comes from the Argument from Remote Desires. In a nutshell, the idea is that the objects of these desires are objectionably remote not because they are remote in space or time, nor because they are not “about our own lives,” nor because they are remote in importance or priority to us; rather, they are objectionably remote to *what we are aware of*. The reason Parfit is not

benefited from the stranger's being cured is that he never hears about it. The reason Kagan is not benefited by the primeness of the number of atoms in the universe is that he doesn't realize that the number of atoms in the universe is prime. For a moment, Griffin even appeared to be taking this line. Let's quote him out of context:

The breadth of the account, which is its attraction, is also its greatest flaw. ... It allows my utility to be determined by things that I am not aware of ... .

He should have stopped there.

Many traditional desire theorists will regard the emendation I am suggesting as treasonous. I am suggesting we incorporate a watered down version of what some philosophers<sup>27</sup> have called the *experience requirement*. "Such a condition would," as L.W. Sumner (1996, p. 127) puts it, "stipulate that a state of affairs can make me better off only if, in one way or another, it enters or affects my experience." Sumner goes on to suggest one sort of desire theory that includes the experience requirement. According to this theory, the states of affairs that are intrinsically good for a subject are states of affairs in which the subject desires some proposition, in which the proposition is true, *and in which the subject is aware that the proposition is true*.<sup>28</sup>

This revision is, I think, enough to avoid the Argument from Remote Desires. Such a theory will not imply that Parfit is made better off by the satisfaction of his desire that the stranger be cured, because Parfit is not aware that the stranger is cured. Nevertheless, I propose a more extreme revision.

---

<sup>27</sup> E.g., Griffin (1986, p. 13).

<sup>28</sup> A theory like this is also alluded to in Brandt (1982, p. 172).

### 4.3.2 Subjective Desire Satisfactionism: The Atoms of Welfare

I propose that we require merely that the subject *believe* that the proposition desired is true. Let's call a state of affairs in which a subject simultaneously desires and believes a single proposition a *subjective desire satisfaction*. A *subjective desire frustration* is a state of affairs in which a subject simultaneously desires and disbelieves a single proposition. The theory I wish to defend – *Subjective Desire Satisfactionism* – mirrors its simpler atomistic ancestor. It begins with the following thesis:

**SDS(i):** Every subjective desire satisfaction is intrinsically good for its subject. Every subjective desire frustration is intrinsically bad for its subject.

Less formally, SDS(i) is saying that every time I think I'm getting something I want to be getting, this is good for me; and every time I think I'm not getting something I want to be getting, this is bad for me. (If I want something that I neither believe nor disbelieve, then no subjective desire satisfaction or frustration occurs, and this is intrinsically neutral for me.) Before laying out the rest of the theory, let's investigate some foundational issues concerning this first thesis of Subjective Desire Satisfactionism.

*a. Desire.* For the purposes of Subjective Desire Satisfactionism, I conceive of desire as *the* fundamental pro-attitude. To desire something is to be for it, to favor it, to be "into" it; metaphorically speaking, it is to give it a mental "thumbs up." I assume this attitude takes propositions as its object, and that one can desire that *p* only if he can conceive of *p*. I assume that desires can be had, in principle, towards any proposition, including ones the desirer believes (this is obviously crucial for Subjective Desire Satisfactionism); ones the desirer disbelieves; ones the desirer knows; ones about the



past, present, or future; ones about no particular time; and ones over which one has no control. For example, I want the Red Sox to win the World Series this year. I consider that proposition and I'm for it; I favor it; I give it a thumbs up; I desire it. And as the Red Sox won the World Series last October, I was into it, favored it, gave it a thumbs up; I wanted it to be happening as I took it to be happening.

I assume that desire comes in degrees: anytime a person desires some proposition, there is some degree, or intensity, to which he desires it. I assume, perhaps somewhat brazenly, that this intensity is, in principle, measurable cardinally and not just ordinally. This assumption is required for the quantitative theses to follow.<sup>29</sup> Despite the Cartesian flavor of these assumptions, I don't not assume that our desires are transparent to us: I assume we can be mistaken about whether we are desiring something, about what we are desiring, and about how strongly we are desiring it.

*b. Occurrent vs. Dispositional Desire.* One question that has received little attention in the welfare literature is to what kind of desire should a desire theorist appeal: *occurrent desire* (roughly, a desire the object of which is currently, in some sense, "before the subject's mind") or *dispositional desire* (desire which would become occurrent if the subject were to think about the proposition in question)? Naive Preferentists, who think the better life is the preferred life, will, I assume, want to appeal to dispositional desires. This is because they may want to say that some life would be better for a person than another even though he doesn't actually take the time

---

<sup>29</sup> Even if this assumption is false, I trust there would be some mathematically sound way to capture the basic idea behind Subjective Desire Satisfactionism – that welfare consists in the subjective satisfaction of desire, and that the stronger the desire subjectively satisfied, the better.

to consider them. He *would* (occurrently) prefer the one to the other if we were to ask him. This would allow Naive Preferentism to enjoy some of the advantages of Ideal Preferentism, which looks to a person's fairly *remote* dispositional desires (the ones he would occurrently have not only if he thought about it but if he had full information, full imaginative acquaintance, etc.).

Subjective Desire Satisfactionism also makes use of the concept of belief. The occurrent belief/dispositional belief question also arises. It takes only a little reflection to see that SDS(i) has any plausibility only if 'desire' and 'belief' are taken in their occurrent sense. For there are indefinitely many dispositional desire-dispositional belief pairs that coincide in us at any moment. For example, each of us, at every moment of our adult lives, dispositionally desires and believes that she will not be killed in three seconds by a falling meteor. But it would be absurd to suppose our lives are continually made better by this fact (not to mention the millions of others like it of which we, and no one, will ever become aware). Furthermore, I see no reason to think there are not *infinitely many* propositions dispositionally desired and believed by each of us at each moment of our lives. If we allow them all to count, we will have a hard time explaining how any two actual lives could differ in well-being; each life would contain infinitely many subjective desire satisfactions and frustrations.

Is it plausible to suppose that I just made your life better by causing you to desire and believe occurrently and simultaneously that you will not be killed in three seconds by a falling meteor? I say Yes. To desire and believe the same thought is, I

say, to think a happy thought.<sup>30</sup> It is to be happy about the thing, or to be pleased about it.<sup>31</sup> Subjective Desire Satisfactionism is a “mental state theory” of welfare: according to it, how well things go for a person depends solely upon her mental states. The goods of life, on this view, are certain mental states; they are good mental states to be in; they are thoughts wherein the way you want the world to be and the way you see the world to be coincide; they are thoughts wherein how you think the world should be is how the world is, as far as you can tell.

*c. Intrinsic vs. Extrinsic Desire.* It is a fairly common idea that normative theories in which desire plays a central role should be restricted to count only *intrinsic* or *basic* desires, and ignore *extrinsic* or *derived* desires. In discussing the Present-aim Theory of rationality, Parfit writes (1984, p. 117):

In deciding [what would best fulfill someone’s present desires], we should ignore derived desires. These are desires for what are means to the fulfillment of other desires. Suppose that I want to go to some library merely so that I can meet some beautiful librarian. If you introduce me to this librarian, I have no desire that is unfulfilled. It is irrelevant that you have not fulfilled my desire to join this library.<sup>32</sup>

For desire theories that employ the dispositional sense of desire, the restriction to intrinsic desire is an option worth considering, for, presumably, for every extrinsic desire D1 there is some at least dispositional intrinsic desire D2 such that D1 is had in virtue of D2. But Subjective Desire Satisfactionism, I said,

---

<sup>30</sup> So also says Wayne Davis (1981).

<sup>31</sup> As I argue in Chapters 5 and 6.

<sup>32</sup> See also Brandt (1979, p. 111).

employs the occurrent sense of desire, and it is not clear that the corresponding principle holds.

That is, it seems possible for a person to have an occurrent desire that we would normally characterize as extrinsic without also occurrently having its associated intrinsic desire. Consider:

*Father and Son.* Father intrinsically wants Son to be respected, to do worthwhile things, and to have a good life. Father thinks if Son gets good grades, then Son is more likely to get these things. So Father strongly desires – but merely extrinsically – that Son get A’s. Father finally sees A’s on Son’s report card, and he is very happy to see them. Father is believing Son has gotten A’s while wanting that Son has gotten A’s.<sup>33</sup>

The spirit of Subjective Desire Satisfactionism is to say that this is a good thing for Father. The way he sees the world to be and the way he wants the world to be match up. However, what if it simply never occurs to Father to consider *why* he wants to see A’s on his son’s report card? Perhaps Father’s focus is on the tasks at hand, not their distant ends; perhaps he has “internalized” the connection between good grades and future success, and rarely gives the connection, or the future success, any further thought. Even so, I don’t think it would be correct to characterize the desire for good grades as intrinsic. For if prompted, Father could, with a little thought, produce his reasons for wanting to see good grades. And he might even insist the good grades are merely a means to this end, even if an end he rarely stops to consider.

---

<sup>33</sup> This example is a variant of one I have heard from Fred Feldman. The problem that this kind of case raises for Subjective Desire Satisfactionism was first brought to my attention by Michael Rubin.

The case of Father and Son might not be terribly rare, either. A person wants and gets a raise, a lottery win, accepted to medical school, his candidate in office, a smile from a beautiful librarian. I want to say each such wanting and (subjectively) getting is good for the person, but I'm not prepared to say that the desire in each case must be intrinsic.

Since the spirit of Subjective Desire Satisfactionism is to say that the subjective satisfaction of these possibly extrinsic desires is good, a natural solution, and the one I accept, is for the theory to count all occurrent desire, whether intrinsic or extrinsic. But might this count too much? I might want to feel the pain of rubbing alcohol in a wound since only then do I think I'll avoid infection, but do I want to say that the desiring and getting of this painful sensation is good? So long as something bad is also happening (which I believe, of necessity, it is, since, in my view, the subject needs an intrinsic desire against feeling the sensation for it to count as painful<sup>34</sup>), it may not be implausible to suppose that the happy thought that one is getting the pain one wants is intrinsically good for one.

To summarize: SDS(i) claims that welfare consists in thinking one is getting what one wants; wants are favorings; the want must be occurrent and can be either intrinsic or extrinsic. To think one is getting what one wants is to be in a mental state that's good to be in.

---

<sup>34</sup> This is defended in Chapter Five.



### 4.3.3 Subjective Desire Satisfactionism: The Value of the Atoms

What determines how good these good thoughts are? That is, What determines the value of a state of affairs in which a person is desiring and believing the same thing? I follow the simple and natural idea that the happier the happy thought, the better. That is, the stronger the desire and the longer its subjective satisfaction lasts, the better it is for its subject. For simplicity of presentation, however, let's assume that desires are very brief entities – so brief that there is no time for their intensity to change. What we would normally describe as a change in intensity of a single desire we will, for the purposes of the theory, describe as an occurrence of a new desire of a different intensity for the same proposition. This enables us to say:

**SDS(ii):** The intrinsic value for its subject of a subjective desire satisfaction = the intensity the desire. The intrinsic value for its subject of a subjective desire frustration = – (the intensity of the desire).

Another foundational issue immediately arises:

*d. Intensity of Desire.* What is meant by 'intensity of desire'? Preferentists who think the better life is the preferred life will likely want to appeal to some sort of dispositional preference ordering conception of intensity of desire. On this dispositional model, there is no appeal to any felt intensity of desire; as a rough-and-ready but highly imperfect heuristic device, we can imagine that the intensity of a desire, on this first conception, is revealed by how much money the subject would, in a moment of calm reflection and vivid awareness of the object of the desire, be willing to spend to have the desire satisfied. Subjective Desire Satisfactionism appeals to a different conception of intensity of desire. I have in mind a non-dispositional, occurrent strength of desire. It need not coincide with the desire's location in the subject's dispositional preference

ranking. However, I do not want to commit myself to the idea that there are always (or even ever) *felt* intensities to desires.<sup>35</sup> Instead, I just assume there is a fact of the matter about how strong some occurrent desire is at some time.

It will not always be easy to tell how strong one's desire for something is. Sometimes it is just immediately obvious to us that our desire for something is very strong, or very mild; sometimes we get an idea of its strength after the fact, by the behavior the desire causes; many times, we will just never know.

It can be that the relative intensity of two desires on this latter (non-dispositional, occurrent) conception is different, at a given time, from that on the former (dispositional preference ordering) conception, at that same time. His team up by one run in the bottom of the ninth with two outs and bases loaded, a baseball fan may have quite an intense desire for a strike out. If the game went to a commercial and we had a chance to calm him down and ask him which he prefers – that the batter strike out, or that his lottery numbers hit – he would surely say the latter. Nevertheless, before the commercial break, his occurrent desire for the strike out had an occurrent intensity far stronger than the occurrent intensity of the desire to win the lottery. The most common way this happens is when we simply are not thinking about the object of the one desire, so that its occurrent intensity is zero. Another way the occurrent intensity can differ from the dispositional intensity is when the desirer is swept up in the heat of the moment; our fan might more strongly desire the strikeout even after we cause him to

---

<sup>35</sup> Though of course I do not deny that we have beliefs about the intensities of our desires, and that sometimes when we believe something, we say that we “feel” that it is true. And I do not deny that very intense desires for things we lack can at least have feelings associated with them.

consider the lottery ticket, if it is before he has a chance to calm down. Yet another way the occurrent intensity can differ from the dispositional intensity is when the object of the one desire is less vividly before the desirer's mind; our fan might have completely calmed down, but might just fail really to consider the prospect of winning the lottery, and so his occurrent desire for it might be less occurrently intense than dispositionally intense.

Thus we have another way in which Subjective Desire Satisfactionism makes a serious departure from traditional desire theories of welfare. In traditional theories, one's rational desires – those had with full information, in a cool moment, while reasoning properly, and while vividly imagining – typically play the important role. Subjective Desire Satisfactionism is anti-rationalist: it appeals to one's actual, not necessarily informed, not necessarily vivid, not necessarily cool, calm, and collected desires.

#### 4.3.4 Subjective Desire Satisfactionism: The Value of a Life

Subjective Desire Satisfactionism determines the value of a complex whole for a person (such as his whole life) in the most straightforward way, by summing over the values of the atoms of value in the complex whole:

**SDS(iii):** The intrinsic value of a life (or life segment) for the one who lives it = the sum of the intrinsic values of all the subjective desire satisfactions and frustrations that occur in that life (or life segment).

It is fashionable to maintain that holistic features of one's life – such as whether it is gradually improving – can make a contribution to its value that is not reducible to the contributions made by the momentary stages of the life. SDS(iii) baldly denies this

idea. On SDS(iii), any value that a holistic feature contributes to a life will have to be contingent and indirect – it will happen only when the liver desires and believes, at individual moments, that her life have some holistic feature.

Subjective Desire Satisfactionism, or SDS, is the conjunction of SDS(i), SDS(ii), and SDS(iii). According to SDS, every time you think you're getting something you want, your life is made better; every time you think you're not getting something you want, your life is made worse. You might be reading a paper on welfare and therefore believe that you are, and while you are reading it you might really be wanting to be reading it, for its insight and originality. If this happens to you, things are going well for you according to SDS, and your life is made better as a result. On the other hand, you might find yourself reading a paper on welfare you think ill-conceived and long-winded; you might therefore strongly wish you weren't wasting your time reading the paper; according to SDS, things aren't going so well for you then.

#### **4.4 Subjective Desire Satisfactionism at Work**

My case on behalf of SDS consists in demonstrating how it handles the many objections to the desire approach to welfare that we have considered so far.

##### **4.4.1 Subjective Desire Satisfactionism and the Arguments from Ignorance, Weakness of Will, and Unimaginativeness**

It is clear that we can desire things it would not be good for us to get. Because this can happen as a result of ignorance and also as a result of unimaginativeness, desire theorists of welfare are drawn to favor one's informed and imaginatively acquainted

counterfactual desires over one's actual ones. I don't know which counterfactual desires they might consult to avoid the parallel objection based on weakness of the will.

I favor an altogether different strategy for responding to these three objections. First, distinguish *intrinsic* and *all-things-considered* value for a person. A state of affairs is *intrinsically good* for someone iff it is good in itself for that person, and not for what else it might accompany; in other words, state of affairs *p* is intrinsically good for a person iff given two possible lives that are exactly the same except with respect to *p*, the *p*-life is better for the person than the not-*p*-life. A state of affairs is *all-things-considered good* for a person iff it makes her life better; in other words, iff had it not occurred, the whole life she would have lived would have been worse for her than her whole actual life.<sup>36</sup> Many states of affairs that are intrinsically neutral (e.g., Bob's buying a lottery ticket) can be all-things-considered good. Even states of affairs that are intrinsically bad can be all-things-considered good – for example, if they lead to intrinsic goods that outweigh their intrinsic badness.

This, I submit, is what happens in the cases upon which arguments from Ignorance, Weakness of the Will, and Unimaginativeness are based. Ignorant Ignacio desires to quench his thirst, and the only way he can do it is by drinking from the river. Unbeknownst to Ignacio, the river is poisoned, and drinking from it will make him terribly ill for weeks. So, (1) it seems it would be bad for him to subjectively satisfy his desire to quench his thirst (for to do this, he'll actually have to satisfy it, we can suppose, and if he actually satisfies this desire, he'll become terribly sick). But (2) SDS seems to imply that it would *not* be bad for Ignacio to subjectively satisfy his desire to

---

<sup>36</sup> Cf. Feldman (1991, 213-214).



quench his thirst (and that in fact it would be good, since all subjective desire satisfactions are good, on SDS). Therefore, SDS is false.

The argument equivocates. To make premise (1) plausible, 'bad' must mean all-things-considered bad. There's nothing inherently damaging about satisfying a desire not to be thirsty.<sup>37</sup> It's bad in this case only for what it leads to. It is all-things-considered bad: it makes Ignacio's life worse than it would have been had he not satisfied that desire. To make premise (2) plausible, 'not bad' must mean not intrinsically bad. SDS is perfectly compatible with the claim that it is all-things-considered bad for Ignacio to subjectively satisfy this desire. In fact, given certain plausible ways to spell out the details of the case, SDS will *entail* that it is all-things-considered bad for Ignacio to subjectively satisfy this desire. It will do so given that, were Ignacio to drink, his subsequent sickness would involve enough subjective desire frustrations, which surely it would.

The replies to the Arguments from Unimaginativeness and Weakness of Will work the same way. Were Maggie to become a pop star, her days would be filled with subjective desire frustration. She would repeatedly see that she is being dogged by the paparazzi while desiring not to be so dogged. This life therefore wouldn't be very good, despite her inability to imagine beforehand how bad it would be. Were weak-willed Willie to succumb to weakness, his future would be filled subjective desire frustration (and we can even suppose he knows it). He would suffer horrible toothache

---

<sup>37</sup> In some of these replies, the subjective element of SDS plays no role, and so, to avoid cumbersome language, I will sometimes leave off the 'subjectively' qualifier.

while desiring not to suffer so. This future therefore wouldn't be very good, despite the fact that he prefers it to the better one in which he visits the dentist.<sup>38</sup>

#### **4.4.2 Subjective Desire Satisfactionism and the Argument from Self-Sacrifice**

SDS's atomism, rather than its subjective element, is what allows it to answer these last three arguments. It also allows it to answer the Argument from Self-Sacrifice. The fully and vividly informed father prefers a world with a much shorter life for himself so that his kids may thrive. But, given SDS, this is compatible with that world being worse for the father. This is because it could be that, were the father to choose against suicide, he would eventually get over his failure to give his kids outstanding lives, and he would go on to enjoy a long life filled with subjective desire satisfaction. This is possible even though the father prefers the shorter life over the longer, subjective-desire-satisfaction-filled life. For there is no necessary connection between an agent's knowing that some possible future maximizes his subjective desire satisfaction and the agent's most preferring that future. We can be motivated by desires other than a desire to think in the future that we will be getting what we want. Therefore, SDS, unlike Ideal Preferentism, is compatible with self-sacrifice.

#### **4.4.3 Subjective Desire Satisfactionism and the Argument from Self-Loathing**

Here is how the Subjective Desire Satisfactionist should describe what goes on in Kraut's self-punishment case. I want to be badly off, and so choose a job that is arduous, boring, and insignificant. Although this is the outcome I want, by no means

---

<sup>38</sup> See chapter 2 for a more detailed presentation and defense of the ideas in this section.

does it follow that, at each of the individual moments in the outcome, I am subjectively getting everything I want. Necessarily, when work is genuinely arduous, one is having desires subjectively frustrated. Likewise, if one feels bored, then one wants to be doing something other than what one is doing. Does it make sense to say that someone found some stretch of time boring, or arduous, but that he was getting everything (as far as he could tell) that he wanted at every moment of that stretch of time?

So each day at this job, I'm miserable. My life is filled with subjective desire frustration. But that's not the whole story, for I do have at least one subjective desire satisfaction on my plate: my simultaneous desire and belief that I be badly off. That's one point in my favor against all the points against me. Nevertheless, I still succeed in punishing myself. As before, it is the atomism of subjective desire satisfactionism rather than its subjective element that renders it compatible with self-punishment.<sup>39</sup>

#### **4.4.4 Subjective Desire Satisfactionism and the Argument from Changing Desires**

SDS is able to answer the Argument from Changing Desires because it requires something akin to *concurrence*: desire and belief must be simultaneous for a subjective desire satisfaction to occur. So if a person's longstanding desire changes, we do not benefit her by making her believe the no-longer-held desire is satisfied, for then she'll have the belief without the desire. More traditional desire satisfaction theories are different in this respect. Since merely past desires about the (then) future can be satisfied, we can, according to these theories, benefit people by giving them what they no longer want. This is what happened in Ellie's 50<sup>th</sup> Birthday Party.

---

<sup>39</sup> See chapter 2, §2.6 above for a more thorough discussion of this response.

Traditional desire theories are both backward- and forward- looking. If you want to know whether some act of yours will benefit your friend, you must know your friend's future desires as well as his past desires with respect to the outcomes of your act. This is why, if a traditional desire theory like Simple Desire Satisfactionism is true, we must consult Ellie's past desires to decide what to do for her.

Subjective Desire Satisfactionism, on the other hand, is forward-looking only. If you want to know whether some act of yours will benefit your friend, you need to know only the impact your act will have on your friend's future desires and beliefs. Past desires are irrelevant. SDS's subjective element ensures this. Since mental states are all that matter on SDS, and since we cannot affect the mental states of the past, no facts about the past play a part in determining well-being. So, if SDS is true, we can ignore Ellie's past desires in deciding what to do for her. If SDS is true, then (filling in the facts of the case in a natural way) it follows that we benefit Ellie by giving her easy listening on her birthday rather than rock 'n' roll. If we give her easy listening, then she will believe that she is getting easy listening while she is wanting easy listening. It will therefore be a great party for her. If, on the other hand, we were to give her rock 'n' roll, then she would believe that she is getting rock 'n' roll while she is wanting not to be getting rock 'n' roll. It would therefore not be a very good party for her.

#### **4.4.5 Subjective Desire Satisfactionism and the Arguments from Unwanted Ideal Lives and Unwanted Ideal Desiderata**

The Argument from Unwanted Ideal Lives was designed to make trouble for desire theories that are holistic and idealistic (or in any way hypothetical). There is an

atomistic counterpart, which we can call the *Argument from Unwanted Ideal Desiderata*. This argument makes trouble for a desire theory that is atomistic and idealistic, one according to which welfare consists in the satisfaction of one's ideal desires (the desires one would have if one were fully and vividly informed). One's ideal self can want things to which one's actual self is strongly averse. In such cases, the target theory implies, unacceptably, that we benefit the actual person by force-feeding him these merely ideally wanted things.

But SDS has no such implication. It is an *actual* desire-satisfaction theory; the satisfaction (subjective or otherwise) of merely ideal desires doesn't count, and so we don't benefit by this sort of force-feeding, according to SDS.

#### **4.4.6 Subjective Desire Satisfactionism and the Argument from Posthumous Harm**

SDS does not imply that events that occur after one's death can alter how good one's life went. This is because, in order for an event to affect one's well-being, it must affect one's beliefs or desires; since the dead have neither, well-being is unalterable after death. Count men happy once they're dead.

Note that SDS does not ignore so-called now-for-then desires.<sup>40</sup> Desires about the past and desires about the future are now-for-then desires. Now-for-now desires are desires about the present. I think it would be implausible for a desire theory to count only now-for-now desires. But SDS makes no such restriction. If a person desires that his body be buried rather than cremated (a now-for-then desire), then this is relevant to welfare, according to SDS. For if the person also believes that his body will be buried

---

<sup>40</sup> So called by Hare (1981, pp. 101-106).



rather than cremated, then he is benefited, according to SDS. On SDS, the desires relevant to welfare can be about any time, past, present, or future, or about no time at all.

#### **4.4.7 Subjective Desire Satisfactionism and the Argument from Remote Desires**

We have already seen how SDS responds to the Argument from Remote Desires. This argument actually provided our main motivation to move to a theory that contains a belief element. This is how we have drawn the line between the remote and non-remote objects of desire. The remote objects are far from our awareness; they are the one's about which we have no beliefs one way or the other.

#### **4.4.8 Subjective Desire Satisfactionism and the Argument from Unworthy Desires**

The arguments featured in these remaining two subsections (4.4.8 and 4.4.9) have been saved for last because the responses to them may be regarded as “bullet-biting.” The Argument from Unworthy Desires strikes at the heart of the desire approach to welfare. The objection contends that it is possible simply to desire the wrong things – and “wrong” in the sense that getting the things damages (or at least fails to enhance) our welfare. Above we considered malicious desires (like those of a sniper terrorizing a community), degrading desires (like those of Moore’s perpetual indulger in bestiality), tasteless desires (like those for Muzak over Mozart), and pointless desires (like those of Rawls’s grass counter).

As I have explained above, I don’t think a desire theorist should be moved by these cases. It does not seem to me at all implausible to maintain that people who

engage in malicious activities, degrading activities, tasteless activities, or pointless activities can have lives that are perfectly good *for them*. This is just what they want, after all. In fact, I think a theory that denies this possibility is itself open to criticism. A conception of well-being according to which malicious, degrading, tasteless, or pointless activities can't enhance seems too moralistic. Well-being is not a measure of your virtue, dignity, or taste. It's a measure of how well things go *for you*.

#### **4.4.9 Subjective Desire Satisfactionism and an Argument from Anti-Paternalism**

If an eccentric billionaire could guarantee you a Brand New Life, along the lines of the one described above, then, so long as we fill in the details properly, SDS implies that it would be better for you to accept this life over the one you would otherwise get and much more strongly prefer. The fact that the Brand New Life on offer is totally alien and unappealing to you at the moment doesn't matter, given SDS. Any subjective desire frustration this might cause initially can easily be outweighed by the benefits the Brand New Life brings.

Quite a few theories of welfare have the same implication. For a theory to have this implication, it just has to be possible for the fundamental source or sources of welfare to vary independently from one's current preferences regarding the Brand New Life and its alternative. So all forms of hedonism have this implication: the Brand New Life on offer could be both alien and filled with pleasure. Objective List Theories have this implication: the Brand New Life on offer could be both alien and filled with items on the Objective List. And atomistic desire theories like SDS also have this

implication: the Brand New Life on offer could be both alien and filled with many instances of desire satisfaction, subjective or otherwise.

It is important to recognize this, for if, like me, you think that holistic forms of the desire theory (the only kind of theory that seems immune to this objection) are implausible, then you will agree that this defect (if it is a defect) is a low tide that lowers all ships. It therefore becomes a dialectically ineffective argument against SDS; it fails to provide a reason to choose some other theory *over* SDS.

How bad is the implication that one is imprudent in rejecting the Brand New Life? Perhaps it is not so crazy to say, “I recognize that this Brand New Life would be a better one for me; nevertheless, I turn it down. It is too alien to me, and I just don’t want to give up my current interests and projects. I prefer a worse life for the sake of my current interests.” I don’t think a theory that makes it coherent to turn down a better life in this way – even when one is not motivated by duty and is instead just thinking of himself – is unacceptable. What may be harder to accept, however, is the claim that such a decision would be irrational.

One is forced into this corner if one accepts SDS along with the idea (embodied in the Self-interest Theory mentioned above) that it is uniquely rational to promote one’s own welfare. Now, I reject this idea, since I think moral considerations provide reasons to do things that may not be in our best interests. But this doesn’t really solve the problem, because in A Brand New Life, your reasons for turning down the better life, we can suppose, are non-moral. To avoid the result that your rejection of the Brand New Life is irrational, we may need it to be that a person’s present desires themselves

provide him with reasons, and that they do this independently of any connection with well-being.

I believe that this needed idea is already fairly common, and so many of us will not be moved by the idea that you have no reason to stick with your ordinary life over the Brand New one. It should be admitted, however, that even if we establish that you have *some* reason to stay with your ordinary life, it could still be that this reason is outweighed by the reasons you have for choosing what's best for yourself, and so there would remain a sense in which your choice is irrational.

#### 4.5 Concluding Remarks

This paper undertakes a major task – the defense of a full-blown theory of welfare. I do not have the space here to consider all the possible objections to Subjective Desire Satisfactionism. So my strategy has been to select a large number of the most influential objections to the desire approach to welfare, and to show how SDS can answer them. I do plan to extend this project into the future, especially by considering arguments targeted specifically towards SDS.

But I will close by responding to one such complaint, one targeted towards the feature that sets SDS most strongly against the desire theoretic tradition: that it is a mental state theory. If including an experience requirement in the theory is enough to answer the Argument from Remote Desires, why also *exclude* a “world requirement” from the theory? Why not say that welfare consists in having desires satisfied *both* subjectively and objectively?

The answer is: because what you don't know can't hurt you. When the world changes without affecting our experience of the world, I don't see how this can count as a harm. And if the quality of a person's experiences improve in that he believe that the world comes to be the way he wants it to be, this seems to me all that is required to benefit the person. Ignorance is bliss; happiness is a state of mind.

Admittedly, this puts Subjective Desire Satisfactionism in a tough dialectical position. It is a desire theory of welfare made for hedonist intuitions. Those inclined towards desire theory may see it as too much of a concession to a mental state approach. Those inclined towards hedonism may see it as a mental state approach that's missing the pleasant zing that makes good mental states good. As for the former complaint, I ask desire theorists to reflect again on what can constitute a benefit or a harm. As for the latter complaint, I think that if we have the happy thought that occurs when desire and belief coincide, then that's actually all there is to getting pleasure.<sup>41</sup> So I prefer to see SDS as a happy reconciliation of the two great monistic traditions in the quest to discover what makes a person's life go well.

---

<sup>41</sup> As I argue in chapter 5.



## CHAPTER 5

### THE REDUCTION OF SENSORY PLEASURE TO DESIRE

#### 5.1 Introduction

##### 5.1.1 Kinds of Connection Between Pleasure and Desire

We all agree that there are some interesting connections between pleasure and desire. For instance, many people, on many occasions, desire pleasure. When a person desires something other than pleasure, it is often because he thinks getting it will lead to his getting pleasure. If a person desires something, the idea of the thing will typically be pleasing to her. If a person knows that getting something would bring him pleasure, he often will form a desire for the thing. There are analogous connections between pain and aversion.

There are bolder claims about the linkage between pleasure and desire, but, naturally, they are more controversial. Many of these bolder claims can be gotten by removing the qualifiers (like ‘often’, ‘typically’) from their weaker cousins in the preceding paragraph. For instance, one claim is that everyone is motivated exclusively by his or her desire for pleasure (and aversion to pain). Sometimes the doctrine of psychological hedonism is put that way.

The most natural way to interpret these bolder claims is as claims of empirical psychology rather than of conceptual necessity. Few psychological hedonists, I think, want to commit themselves to the idea that psychological hedonism is a necessary

truth.<sup>1</sup> They think rather that it is a mere contingent matter of psychological fact about human beings, a claim to be established in the laboratory rather than in the armchair.

This suggests a third class of alleged connections between pleasure and desire: those that do affirm a conceptual, or analytic, or metaphysically necessary connection between the two phenomena. In this paper, my focus will be on claims in this third category. My larger project is to develop and defend a set of theses to the effect that pleasure is reducible to desire. But in this paper, I will limit my focus to one kind of pleasure: sensory pleasure.<sup>2</sup>

The idea that pleasure is explainable in terms of desire is by no means new – sometimes it goes by the name ‘the motivational theory of pleasure’. What I hope to add to the debate are more carefully formulated versions of the reductive theses – many formulations of the alleged conceptual connection between pleasure and desire are less than totally clear, other initially plausible formulations are rightly rejected – and more sustained defenses of them, by responding to objections.

### 5.1.2 A Theory of Pleasure, and Why

The reduction of pleasure to desire that I favor is a *theory of pleasure*, a thesis (or set of theses) that purport to reveal the nature of pleasure. The theory is reductive in the following sense: if it is true, then facts about pleasure just are facts about desire; we

---

<sup>1</sup> Perhaps Mill (1998, p. 85) is one exception: “to desire anything, except in proportion as the idea of it is pleasant, is a physical and metaphysical impossibility.”

<sup>2</sup> I will be discussing pain as well, and I endorse the view that pain is reducible to desire in similar fashion. For stylistic simplicity, however, many remarks to come mention pleasure only, with the analogous claim about pain left implicit.

could remove the word ‘pleasure’ and its cognates from English without diminishing its expressive power.

I find the attempt to discover the nature of pleasure a worthwhile project in its own right, no less interesting than attempts by philosophers to discover the nature of other important phenomena, like knowledge, freedom, causation, value. But there are other reasons to be interested.

a. The concept of pleasure plays an important role in many areas of normative ethics: axiology, welfare, rationality, moral obligation. We cannot hope to understand fully many important theses in these areas of philosophy until we understand what pleasure is.

b. There are two puzzles about pleasure – the heterogeneity problem and the oppositeness problem (both introduced below) – that a theory of pleasure should solve. So a good theory of pleasure will remove some mystery from the world.

c. Below I demarcate three categories of pleasure. At least at first glance, they appear to be three irreducibly distinct kinds of pleasure. Perhaps a theory of pleasure can unify these categories after all. The complete reduction of pleasure to desire I defend (only part of which, as I said, I can discuss here) is supposed to do exactly this.

d. The theory of pleasure I endorse has another interesting implication: it implies that a form of axiological hedonism is reducible to a version of the desire theory of value. Two of the leading competitors in the theory of value are not such great rivals after all.<sup>3</sup>

---

<sup>3</sup> I defend this in chapter 6.

### 5.1.3 Desire

For the purposes of this paper, I take the concept of desire as primitive. I offer no account of what desires are. Nevertheless, let me say a few things about desire. There seem to be different senses of the word ‘desire’ – I want to be sure we have the right one before our minds.

Some philosophers distinguish two senses of ‘desire’, a narrow sense and a wide sense, distinguished by means of a relation to behavior.<sup>4</sup> According to the narrower conception, it is possible to perform some action intentionally, voluntarily, and knowingly even though one has no desire to perform it. On this narrow sense, we might be motivated not by our own desires but by the desires of others, or by what morality or prudence requires. We can say, “I didn’t want to do it, but I felt I had to do it.” On the wider notion of desire, this is not possible. If some act is truly intentional and voluntary, then *ipso facto* its agent had a desire to perform it.

My aim is the less ambitious task of reducing pleasure to the wider notion of desire. We can ignore the narrow sense of ‘desire’ for the purposes of this paper. In fact, it seems to me that I want to work with what is probably the widest sense of the term ‘desire’. As I understand the notion to be used here, desires are some kind of “pro-attitude” – *the fundamental* pro-attitude, I would want to say. To desire something is to favor it, to be for it, to be “into” it. Metaphorically speaking, it is to give the thing a mental “thumbs up.”

I will assume that the objects of this attitude are propositions, so that the “fundamental locution” is the following:

---

<sup>4</sup> See, e.g., Schueler (1995), pp. 1-7.

*Subject S desires to degree n at time t that proposition p be true.*

All elliptical and other colloquial attributions of desire – including those making use of the word ‘want’ – can, according to this assumption, be translated into this canonical form. I assume that we can, in principle, take this single attitude towards any sort of proposition, whether past, present, or future; true or false; or believed or disbelieved.<sup>5</sup>

## 5.2 The Heterogeneity Problem

As has been widely observed, pleasure is a diverse and varied phenomenon. There are bodily pleasures, like those gotten from relaxing in a Jacuzzi tub, from sunbathing on a warm beach, or from sexual activities. There are gustatory and olfactory pleasures (maybe they, too, qualify as “bodily”). There are what we might call “emotional pleasures,” such as the elation of receiving an ovation or the prideful satisfaction of completing a difficult and worthwhile project. There are more “cognitive” pleasures, such as the pleasure derived from working on a crossword puzzle, from reading an insightful philosophy paper, or from listening to an amusing anecdote. There are aesthetic pleasures, like those derived from listening to beautiful music or from taking in a powerful sculpture. These are all pleasures. Why? What is it about each of these things in virtue of which it is correctly classified as a pleasure? This is “the heterogeneity problem.”<sup>6</sup>

---

<sup>5</sup> This assumption/stipulation is the subject of a coming objection (§4.6.1).

<sup>6</sup> Feldman (1997, p. 84) uses a similar expression, as does Carson (2000, p. 14). Brandt (1979, p. 37) and Korsgaard (1996, p. 148) also find heterogeneity problematic.



There is a heterogeneity problem for pain as well. There are bodily pains; perhaps there are gustatory and olfactory pains<sup>7</sup>; there are definitely emotional and cognitive pains. What is it about each of these things in virtue of which it is correctly classified as a pain?

A nice, clean answer to the heterogeneity problem, and one that might come first to mind, is that pleasure is some one kind of immediately felt quality, and that various human activities (like eating, reading, listening to music) cause us to feel it. This is the *Felt-Quality Theory*.<sup>8</sup> This answer actually denies that pleasure itself is heterogeneous – just its causes are. There are well-known arguments against Felt-Quality Theories, and, suffice it to say, the phenomenology just doesn't bear it out – there doesn't seem to be any one feeling (or even “hedonic tone”) common to all occasions on which we experience pleasure or enjoyment.<sup>9</sup>

Another popular solution to the heterogeneity problem is that, in each case, the subject *desires* the pleasure. He desires to be tasting the taste of the food, to be reading

---

<sup>7</sup> Or at least *unpleasant* gustatory and olfactory sensations. More to come (§6.5) on the distinction between painful and merely unpleasant sensations.

<sup>8</sup> Advocates of some form of the Felt-Quality Theory (there are at least two distinct varieties) include James Mill (1869, ii p. 184), G.E. Moore (1993, p. 64), C.D. Broad (1930, pp. 229-231), Karl Dunker (1941), and Timothy Sprigge (1987). Sumner (1996, p. 89) classifies James Mill this way. Both Sumner (1996, pp. 88-89) and Sobel (2002, p. 240, 2n) seem to regard Bentham as belonging in this category, but it is not clear to me that he belongs. I follow Carson (2000, p. 13-14) in using the name ‘Felt-Quality Theory’ to cover both the *Distinctive Feeling View* and the *Hedonic Tone Theory* (so-called by Feldman (2001, p. 663)).

<sup>9</sup> Alston (1967), primarily following Ryle, presents a series of arguments against a Felt-Quality View. Many philosophers make the point about phenomenology, e.g., Sidgwick (1907, p. 127), Brandt (1979, pp. 37-38), Feldman (1997, p. 87), Carson (2000, p. 14), and Sobel (2002, p. 241).

the paper, to be listening to the music. What makes each of these experiences pleasurable is that the subject wants to be having it. People who entertain this hypothesis are on to a desire theory of pleasure.<sup>10</sup>

### 5.3 The Desire Approach to Pleasure and “the Oppositeness Problem”

Among those people are some philosophers. After rejecting a preliminary version, William Alston (1967, p. 345) discusses what he sees as the “most promising” formulation of the motivational theory of pleasure:

To get pleasure is to have an experience which, as of the moment, one would rather have than not have, on the basis of its felt quality, apart from any further considerations regarding consequences.

Richard Brandt has endorsed some kind of a motivational theory of pleasure in many places. He describes (1979, p. 38) the “motivational theoretical-construct theory” as follows:

The theory to which we come is, roughly, that for an experience to be pleasant is for it to make the person want its continuation.

This conception isn’t quite Alston’s. Alston’s makes no mention of wanting its continuation (a future-directed desire). Alston speaks of preferring the experience “as of the moment” (a present-directed desire).

---

<sup>10</sup> A nominally similar solution to the heterogeneity problem, which I cannot discuss here, unifies pleasure not through desire but through *desirability*. This is Sidgwick’s view (1907, p. 127): “when I reflect on the notion of pleasure ... the only common quality that I can find in the feelings so designated seems to be that relation to desire and volition expressed by the general term ‘desirable’ ... . I propose therefore to define Pleasure ... as a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable ... .” Rachels (2000, p. 198) speaks favorably of other normative analyses of pleasure and pain.

An immediate advantage of solving the heterogeneity problem by appeal to desire is that it automatically explains a second fact about pleasure – viz., that pleasure is, in some important sense, the *opposite* of pain. The Felt-Quality Theory of pleasure and pain described above would leave it mysterious why, and in what sense, pleasure and pain are opposites. Many pairs of felt qualities (e.g., a sensation of middle C on a piano and a sensation of F# on a banjo) are in no way opposites. But if the felt quality theory is true, then some such pairs are opposites. How could that be? What could make one sensation the opposite of another sensation?

On a complete motivational theory (one about both pleasure and pain), the oppositeness of pleasure and pain is explained. Pleasure and pain are opposites because pleasure is explained in terms of desire, pain is explained in terms of aversion (or desiring not), and desire and aversion are opposites. And if aversion really is just desiring not, as I assume it is, then the oppositeness of desire and aversion is, in turn, explained in terms of the oppositeness of a proposition and its negation.

Derek Parfit (1984, p. 493), another philosopher who wants to solve the heterogeneity problem by appeal to desire, seems to be on to this further advantage as well in the first sentence of the following passage (the rest of the passage contains his endorsement of a desire theory of pleasure):

What pains and pleasure have in common are their relations to our desires. On the use of ‘pain’ which has rational and moral significance, all pains are when experienced unwanted, and a pain is worse or greater the more it is unwanted. Similarly, all pleasures are when experienced wanted, and they are better or greater the more they are wanted.

Thomas Carson (2000, p. 13) states and endorses a reduction of pleasure to desire in the following passage:

The motivational theory of pleasure is the view that the pleasantness or unpleasantness of an experience is a function of one's desires with respect to it qua feeling. A pleasant experience is an experience that one prefers to have rather than not have (abstracting from all considerations about its consequences and preconditions); an unpleasant experience is an experience that one prefers not to have (abstracting from all considerations about its consequences and preconditions). ... The motivational theory provides a plausible account of what all of the many different kinds of experiences we call pleasures have in common.

Other philosophers have considered and/or endorsed other sorts of desire theory of pleasure and pain.<sup>11</sup>

Some of these passages are clearer than others, but we would do well to settle certain things. Brandt's theory makes use of future-directed desire whereas Alston's and Parfit's makes use of present-directed desire. Parfit's theory is about pain whereas Carson's theory is about unpleasantness. Alston's mentions disregarding consequences; Carson's, consequences and preconditions. Brandt's and Parfit's make no such restrictions. Brandt's formulation, but none of the others, requires a causal connection between experience and desire.

#### **5.4 Three Categories of Pleasure**

Divide and conquer. If we divide the heterogeneous class of pleasures into subcategories, we can conquer the heterogeneity problem. To begin, it seems clear that there are sensations, or feelings, of pleasure. If you're like me, you continually experience sensations, and some such sensations you would not hesitate to describe as

---

<sup>11</sup> E.g., Spencer (1871, §125), Edwards (1979, pp. 92-96), Kagan (1992, pp. 173-74), Korsgaard (1996, pp. 147, 148).

pleasant. If you've ever, say, gotten a massage, you know about pleasant sensations. So *sensory pleasure* is one kind of pleasure.

Expressions containing 'pleasure' and its cognates are also used to express a propositional attitude, as when we say that a person is pleased that something is the case, or takes pleasure in some state of affairs. For instance, we can say, "Mary took great pleasure in the fact that her presentation was well received," or "I'm so pleased the Red Sox won the World Series." So *propositional pleasure* is another kind of pleasure. Propositional pleasure is pretty clearly not reducible to sensory pleasure (as a Felt-Quality Theory might require): I, for one, really am pleased the Red Sox won – I'm pleased right now – but I can testify to the fact that I am feeling no pleasant sensations right now.

Yet another kind of pleasure is picked out by the word 'enjoys'. A person can be enjoying a state of affairs at a time only if the state of affairs is occurring at that time. Not so with propositional pleasure. It can be that Stuart is pleased that he will soon be home. But it can't be that Stuart is enjoying that he will soon be home (though he may, I suppose, be enjoying thinking about being home). The enjoyment will come when he finally gets home, and begins enjoying being home. It therefore seems that *enjoyment* constitutes a third category of pleasure.<sup>12</sup>

My focus in this paper is on the first category only: sensory pleasure. Although I want to defend reductions of propositional pleasure and enjoyment to desire as well,

---

<sup>12</sup> Davis (1982, p. 244) discusses further differences between enjoyment and propositional pleasure.



and thereby unify the three categories, the reduction of sensory pleasure to desire is a topic big enough to occupy us here.

### 5.5 The Reduction of Sensory Pleasure to Desire

For there is a heterogeneity problem even just for sensory pleasure. The sensations of pressure felt during a massage, the sensation of warmth had when sunbathing, the taste of a freshly cracked beer, the olfactory sensations experienced when smelling freshly baked bread – each of these is (typically) a sensation of pleasure – that is, each of these is (typically) a sensory pleasure, a sensation that is pleasant or enjoyable. But, as before, why? What is it about each of these sensations in virtue of which it is a sensation of pleasure?

The motivational theory of sensory pleasure provides an answer to the heterogeneity problem for sensory pleasure. The basic idea is that a sensation gets to be a sensation of pleasure if the subject of the sensation desires to be feeling it. Notice that whereas a Felt-Quality solution to the heterogeneity problem attempts to unify sensory pleasure by appeal of some intrinsic feature of the pleasure sensation, the desire-based solution appeals to an extrinsic feature of the sensation: that its subject has some attitude towards it.

To make this basic idea more precise, let's assume that there are such things as *sensory properties*, or *phenomenal properties* – properties such that “there is something it is like” to instantiate them.<sup>13</sup> Let's assume that to experience a sensation is to

---

<sup>13</sup> ‘Phenomenal property’ may be a better term here than ‘sensory property’ because the latter may suggest that the properties in question must be associated with one of the five

instantiate one of these properties. This enables us to identify sensations themselves with events or states of affairs in which some subject instantiates some sensory property at some time. If we let ‘S’ range over subjects of experience, ‘P’ over phenomenal properties, and ‘t’ over times, we can state our first approximation of a reduction of sensory pleasure to desire:

**RSPD1:** a sensation *S*’s instantiating *P* at *t* is a sensory pleasure =df.  
S desires S’s instantiating *P* at *t* .

RSPD1 is an analysis of the property *being a sensory pleasure*, a property had by sensations. It says that this property is had by a sensation just in case the subject of the sensation desires the sensation.<sup>14</sup>

But recall the restriction Alston and Carson make. On Alston’s formulation, “To get pleasure is to have an experience which ... one would rather have than not have ... *apart from any further considerations regarding consequences*” (emphasis added). Why has Alston included this restriction?

He includes it because we can desire sensations that are clearly not pleasant – and even painful – if we want them *for some reason*, if they are means to some desired end. Suppose I’ve got a fresh lesion. I don’t want it to become infected, so I clean the wound with rubbing alcohol. We all know how that feels. But I’m actually *wanting* to feel the sting because I know that only then is the alcohol doing its job. So there’s a very painful sensation – the stinging sensation of alcohol on a fresh wound – that I want

---

sense modalities. But this isn’t so: the property of feeling hungry, for example, would be included in the category I mean.

<sup>14</sup> If the states of affairs that are sensations are not propositions and are instead something like concrete events, then, to better conform to canonical form, we can replace the right-hand side with: ‘S desires that *S*’s instantiating *P* at *t* occur’.

to be feeling. Eventually I get it. RSDP1 therefore implies, unacceptably, that the sensation is a sensory pleasure.

But I don't want to feel the stinging sensation for its own sake. I want to feel it for a reason – namely, that it indicates that I won't get an infection. The theory therefore needs to be restricted in a familiar way: to count only desires for sensations “for their own sake.” This is what Alston was getting at with “apart from any further considerations regarding consequences.” Alston was restricting the view count only what we can call “intrinsic desires” – desires for things for their own sake.

You have an intrinsic desire for something when you just want it – when there is no reason you can give for wanting it, no further thing you want that you think it will bring you, no end to which it is a means. You just want it – for its own sake. When there *is* a reason you want something, then you have an extrinsic desire for it. You want the thing because you think your getting it will, in some way or other, make it likely that you will get something else you want.

Here's a familiar kind of illustration. I want to walk into the pub. Why do I want to walk into the pub? To order a beer. Why do I want to order a beer? So I can drink a beer. Why do I want to drink a beer? So I can taste a beer. Why do I want to taste a beer? Um ... uh ... no reason, I guess. I just do. I want to taste the beer for its own sake, not for anything it will lead to. This last desire is intrinsic, all the others extrinsic.<sup>15</sup>

---

<sup>15</sup> Two further points about the distinction between intrinsic and extrinsic desires. First, the issue is complicated by the fact that we can have both an intrinsic and an extrinsic desire for one and the same thing. Second, ‘extrinsic desire’ is a better term than ‘instrumental desire’. The latter term suggests that the connection between the object of

If the theory is restricted in this way, as

**RSPD2:** a sensation *S*'s *instantiating P at t* is a sensory pleasure =df.  
S intrinsically desires that S instantiates P at t,

it no longer implies that the burning sensation of the rubbing alcohol is pleasant, for my desire for it is merely extrinsic. I want to feel it only because it indicates I will get something else I want.

But RSPD2 is still not quite right. The main problem has to do with time.

Suppose that as a child I enjoyed eating Froot Loops, a sugary cereal. Now I want to taste that same taste of Froot Loops again, the taste I used to love. My desire for this taste sensation is intrinsic – I want it for its own sake. Suppose I hunt down some Froot Loops and get exactly the taste sensation I have been wanting. But suppose it tastes awful – way too sweet! I've outgrown my sweet tooth. It is not clear what RSPD2 implies in this case of changing desires. It is not unreasonable to interpret RSPD2 in such a way that it implies that my taste of the Froot Loops is a sensation of pleasure, for I did intrinsically desire it, at least before I got it.

If Sidgwick (1907, p. 110) were here, he would say I had gotten a "Dead Sea apple." He knew that

... what is desired ... may turn out a 'Dead Sea apple', mere dust  
and ashes in the eating.

Cases like these are *disappointments*. There are also *pleasant surprises*. Sidgwick (1907, p. 45) also knew these:

---

an extrinsic desire and the object of its associated intrinsic desire is always *causal*. But it need not be causal: it can instead be *preventative*, *indicative*, or *exemplifying*.

... many pleasures,—especially those of sight, hearing and smell, together with many emotional pleasures,—occur to me without any perceptible relation to previous desires.

In a similar vein, David Perry (1967, pp. 204-205) writes:

To enjoy a thing it is not logically necessary that one should have desired it or that one would desire its preservation or that one should desire the continuation or repetition of his experience of it. ... I might, on passing a garden by chance, enjoy the scent of flowers without sniffing, lingering, returning, or trying to do these things or having the least inclination to do these things.<sup>16</sup>

To answer the objections from disappointments and from pleasant surprises, we need to refine RSPD2 so that the only intrinsic desires relevant to whether a sensation is a sensory pleasure are those had *at the same time* the sensation is had. Sensation and desire must overlap, as follows:

**RSPD3:** a sensation *S*'s *instantiating P at t* is a sensory pleasure =df.  
S intrinsically desires at t that S instantiates P at t.

RSPD3 can handle disappointment. There was a time I intrinsically desired to be having the taste sensation of Froot Loops. And there was a time at which I had the sensation. But these times never did overlap. As soon as I got the sensation, I no longer wanted it. RSPD3 therefore does not imply that the taste was pleasant. Likewise for pleasant surprises. Perry – I say – forms a desire to be experiencing the sensation the instant he get it. He strolls by, experiences an olfactory sensation, and gives it a “thumbs up.” He can do this without giving a thumbs up (or any thought at all) to the

---

<sup>16</sup> Sobel (2002, p. 253) makes a similar objection. Note that if this objection is good – and I think it is – then Brandt's formulation, which requires a desire for the *continuation* of the experience, is in trouble.



prospect of the sensation's continuation. RSPD3 requires only that he be into it while he is getting it. And of course he is.<sup>17</sup>

RSPD3 is a reduction of sensory pleasure to desire. But it is not saying that sensory pleasures *are* desires, or any other kind of intentional state. Rather, it says that a sensation gets to be a sensation in virtue of its subject having a desire towards the sensation. So RSPD3 does not commit us to any reductionist view about sensations themselves – just about the property *being a sensory pleasure*.<sup>18</sup>

## 5.6 Arguments Against the Reduction of Sensory Pleasure to Desire

### 5.6.1 Arguments from the Concept of Desire

RSPD3 does not require that the subject of a sensation of pleasure desire it before it occurs. He just has to desire (intrinsically) to be feeling it while he is feeling it. But some philosophers have claimed that this can never happen, because we cannot want what we already have. Aquinas held this view (*Summa Theologica*, Ia IIae, 30, 2 ad 1), and his endorsement of it is cited approvingly by Anthony Kenny (1963, pp. 115-116):

Aquinas pointed out that it is as impossible to want what one already has as to remember what is now happening. ... This obvious condition is something strangely neglected by philosophers.

---

<sup>17</sup> Edwards (1979, pp. 95-96) makes a similar reply to a similar objection.

<sup>18</sup> RSPD3, while an “externalist attitude” view (on Sumner’s (1996, p. 91) terminology), is different from another kind of externalist attitude view – the kind that reduces sensory pleasure not to desire but to *liking* or *propositional pleasure* (see, e.g., Hall (1989) and Feldman (1997) this kind of view). Desiring is different from liking since only the latter entails believing its object.

The idea seems to be that it is a simple conceptual truth about desire that one cannot want what one already has. If someone has something, one would be guilty of conceptual confusion to think that he might also want the thing. Let's call the putative conceptual truth about desire

**Falsity:** Necessarily, if S desires that p be true, then p is not true.

If Falsity is true, RSPD3 becomes unacceptable, for it follows from the combination of these views that that there is no such thing as sensory pleasure.

L.W. Sumner (1996, p. 128-130) advances a similar claim about the concept of desire, and one that, in conjunction with RSPD3, yields the same unacceptable result. Sumner maintains that desire is essentially "prospective" – i.e., that we can have desires only about the future. If a person desires at some time t that some proposition be true, Sumner claims, then the proposition must be about a time later than t. Says Sumner (1996, p. 129):

I can desire now only that something occur later. Desires are always directed on the future, never on the past or present. ... In being future-directed in this way, wanting once again contrasts with liking or enjoying. I can (occurently) enjoy only what I already have, while I can want only what I have not yet got.

Let's call Sumner's thesis

**Prospectivity:** Necessarily, if a person S desires at t that p be true, then p is about a time later than t.

My response to both objections is to deny the principle. I think there are clear cases of desires towards true propositions and desires towards non-future propositions. But before looking at any such cases, let's take notice of the fact that the dispute over Falsity and Prospectivity is more verbal than ontological. For surely there is *some* pro-

attitude that we can bear towards things we have (i.e., towards true propositions) as well as towards present and past states of affairs. We can favor, or be into, or give a mental thumbs up to, past and present states of affairs just as well as to possible future ones.

And we can give a thumbs up towards things we have. An attitude is there to be named. The question is simply whether it is ever appropriate to use ‘desire’ or ‘want’ to express this attitude. To ensure that my paper deserves its name, all I need is for there to be *some* sense of ‘desire’ and ‘want’ according to which these conceptual theses are false.

I think one simple example suffices to demonstrate that there is such a sense. Suppose Cheapskate’s car is parked outside. It begins to rain. Worrywart notices, and says to Cheapskate,

“I bet you prefer that your car be in the garage right now.”

But Cheapskate’s car is dirty. He thinks letting it sit in the rain is a cheap way to get it clean. So he replies,

“No, I want my car to be right where it is.”

Cheapskate is expressing a desire, and I think what he says can be literally true. He really can want his car to be right where it is. The object of Cheapskate’s want is *that his car be where it is*. But, of course, Cheapskate’s car *is* right where it is – the object of his desire is true. So Falsity is false.<sup>19</sup> The state of affairs *that Cheapskate’s car be where it is* is also a present state of affairs. Cheapskate wants *now* that he car be where it is *now*. So we have also a case of a non-prospective desire – a desire towards a state

---

<sup>19</sup> This point, or a very similar one, is made also by Matthews and Cohen (1967) against Kenny and Aquinas.

of affairs not about the future but about the present. This one case refutes both principles.<sup>20</sup>

### 5.6.2 An Argument from Desired Non-Pleasures: The Dizziness Case

Fred Feldman (1997) considers and rejects a theory of sensory pleasure like RSPD3. Feldman maintains that it's possible to desire intrinsically to be feeling some sensation without the sensation being a sensation of pleasure. In support of this former claim, Feldman presents a case involving the feeling of dizziness. The case is used first – successfully, in my opinion – to undercut a theory that fails to incorporate an intrinsicality restriction. In the case, a research scientist wishes to understand the phenomenon of dizziness and so performs experiments on himself. He gets himself dizzy and attends to the feeling, wanting to understand its nature. When he feels sensations of dizziness, he wants to be feeling those sensations. If you told him you could make the feelings stop by giving him some Dramamine, he would turn down your offer. For he wants to be feeling them. Nevertheless, the sensations might be thoroughly unpleasant.

In this case, the researcher's desire for the dizziness is merely extrinsic: he desires his feelings of dizziness only because he desires to understand dizziness, and he thinks having the feelings will lead to his understanding them. Therefore, the case refutes only a naïve motivational theory (like RSPD1). But Feldman extends the

---

<sup>20</sup> Parfit (1984, pp. 170-172) presents further considerations in favor of the legitimacy of a wide notion of desire, such as the one employed in RSPD3.

dizziness case. He takes a cue from the phenomenon, exemplified classically by misers, of strong and lasting extrinsic desires turning into intrinsic desires<sup>21</sup>:

As time goes by, I [the researcher] begin to lose interest in my research, and I become more interested in the dizziness for its own sake. On some occasions I spin around to make myself dizzy even though I've long since given up my research project. Dizziness, which I formerly desired only extrinsically, is now something I desire at least partially intrinsically. Nevertheless, the feeling of dizziness, when I get it, may fail to be a sensory pleasure. (Feldman 1997, p. 90)

I think the case is inconclusive. It is not so clear that it's possible for the feeling of dizziness in this case to fail to be pleasurable. I think we can cast doubt on Feldman's contention if we attend to the concept of *liking*, and related concepts, and see how they can be exploited by the reductionist to question the possibility of intrinsically desiring some sensation without finding it pleasant.

Let's suppose that there is just one feeling of dizziness and call this sensory property 'D'. The researcher, whom we will call 'r', instantiates D at some time (call it 't') and thereby experiences a sensation of dizziness, *r's instantiating D at t*. The researcher experiences *r's instantiating D at t*, and as he experiences it, he intrinsically wants to be experiencing it. Now imagine the following interrogation:

Interrogator: "Researcher, you say you *intrinsically desired* at t your feeling dizzy at that time."

Researcher: "That's correct."

Interrogator: "This means that at t you *avored*, or gave a mental thumbs up to, your feeling dizzy at that time."

Researcher: "That's right; that's just what it was for me to intrinsically desire the sensation."

Interrogator: "Let me ask you some further questions. Were you *attracted* to the sensation itself?"

---

<sup>21</sup> See Mill (1998, p. 83).



- Researcher: "I think I should have to say I was. Otherwise, how could it be true that I favored it? If I were to deny an attraction to the sensation, I think I would *ipso facto* be denying an intrinsic desire for the sensation."
- Interrogator: But you can't give a reason you were attracted to the feeling, right? You just were. If you gave some reason for being attracted to it (such as that you found the sensation interesting, or that it would help you with your research), then this would reveal your desire to be extrinsic.
- Researcher: Agreed. I was just attracted to the feeling. And that's that.
- Interrogator: So it seems you would have to say that you *liked* feeling the sensation?
- Researcher: I think I must say that I liked it. After all, I was in favor of it, I would give it a thumbs up, I was attracted to it. Does it make any sense to accept all of this but then to deny that I liked feeling the sensation (for its own sake)?
- Interrogator: So you liked that you were feeling the sensation. Did you *enjoy* or *take pleasure in* the dizziness sensation?
- Researcher: Well I said I liked it, didn't I? Doesn't that just mean the same as that I took pleasure in it.
- Interrogator: Well then I think we have established that you found the sensation pleasant.

In this response to the dizziness case, we go on a conceptual slide from desire to pleasure. The theory says that intrinsically desiring some sensation entails its being a sensory pleasure. We establish this less-than-obvious entailment with more obvious ones. Intrinsically desiring it entails being into it, which entails being attracted to it, which entails liking it, which entails finding it pleasant, which entails its being a sensory pleasure.

I therefore maintain that the researcher's feeling of dizziness is pleasurable. But this admittedly does sound odd. I know what dizziness feels like and it is definitely not pleasant. How could such a feeling ever count as a sensation of pleasure?

Consideration of the diversity of tastes in food can answer this question. It should teach us to be wary of jumping to conclusions about the status of certain sensations in others. I love spicy food. I feel certain sensations of spiciness and I really enjoy them. My mother, on the other hand, can't take it. She finds spicy food literally painful; she wants no part of it. But surely we would balk at any attempt on my mother's part to conclude that my sensations of spiciness are not pleasant. Surely the following speech by my mother would be off the mark as a refutation of RSPD3:

You've sure got a silly view about desire and pleasure, my boy. Remember when you forced me to try your vindaloo? Well I remember it, and the experience was terrible. Those sensations were sensory pains, not pleasures. But you love vindaloo. You really enjoy having those sensations. You desire them for their own sake whenever you get them. Your RSPD3 therefore implies that they were pleasant. But this is obviously false. I've felt those sensations, and I'm certain about one thing: they are not pleasant!

Obviously, no defender of RSPD3 would or should be moved by such an objection. My mother even admits that I enjoy the sensations. She rejects the idea that they are pleasures simply because she has felt one's just like them and didn't enjoy them.

I think Feldman's case is trading, though in a less crude way, on similar prejudices. If his case has some appeal to you, perhaps it is because you, like me, could never find feelings of dizziness pleasant. You might therefore be tempted to conclude that they mustn't be pleasant in the researcher either. But this is just the kind of prejudice my mother exhibits in her faulty argument. It's easier to notice there, because we already recognize great diversity in food tastes. Since there happens to be much less diversity of our tastes about feelings of dizziness, I think we are less inclined to take notice of the prejudice. It is surely bizarre for someone to find dizziness pleasant. But

no more bizarre, I submit, than to have an intrinsic desire to feel dizzy. Present a bizarre case, get a bizarre result.

### 5.6.3 A Second Argument from Desired Non-Pleasures: “Interesting” Sensations

It can happen that we want to be feeling a certain sensation, and that we want to be feeling it in virtue of something about its intrinsic nature, but if asked what it is about its intrinsic nature that attracts us, we will say not that we find it pleasant, but that we find it *interesting*. If we desire the sensation for its intrinsic nature, it seems our desire may qualify as intrinsic. But if what’s attractive about the sensation’s intrinsic nature is merely that it is interesting, it may be that the sensation is not a sensory pleasure. But if so, then we have a case of an intrinsically desired but non-pleasant sensation, and so a counterexample to RSPD3.<sup>22</sup>

An objection along these lines is actually made against Feldman’s own theory of sensory pleasure by L.W. Sumner (1998). Sumner (pp. 177-78) asks you to

Think of a feeling (say of “pins and needles” in the legs) which you find intrinsically interesting and of being pleased that you are having it on a particular occasion because of its interesting intrinsic qualities; this does not seem sufficient to make it pleasurable.

The objection is slightly different, since it is against a different theory (one, incidentally, that reduces sensory pleasure to propositional pleasure), but the basic idea is the same: we can want to be (and even enjoy) feeling some sensation for its own sake without the sensation’s qualifying as a sensory pleasure.

---

<sup>22</sup> Though I didn’t do so above, Feldman’s dizziness case could actually be interpreted as an instance of this objection.

My response to the “interesting sensation objection” depends crucially the intrinsic desire/extrinsic desire distinction. In a word, desires for sensations that are had in virtue of the sensation’s interestingness are not intrinsic desires. The reply to (the modification of) Sumner’s case, therefore, is that the desire to be feeling the pins and needles sensation, despite its having to do with features intrinsic to the sensation, is an extrinsic desire. RSPD3 therefore does not imply that the sensation is a sensory pleasure.

Recall what we said above about intrinsic desire. When a person has an intrinsic desire, there is no reason he wants the thing intrinsically desire – he just wants it.<sup>23</sup> He can’t say he wants it because it will bring about this, or prevent that, or indicate this, or be an instance of that. If he could say that, then his desire would be extrinsic.

Now, suppose we look at you in disbelief. We can’t believe you want to be feeling the pins and needles sensation, so we ask you to explain. *Why* would you ever want to be feeling that pins and needles sensation? You reply: because the feeling is interesting to me. But this is crucial. You have given a reason for your desire! This is evidence that your desire is extrinsic. You don’t want those sensations just for themselves. You want them only because they are interesting. If they ceased being interesting, or if you ceased wanting something interesting, you would lose your desire to be feeling the sensation. This shows that the desire for the sensation is extrinsic.

---

<sup>23</sup> This, of course, is not to deny that there is a causal explanation for the existence of the desire – of course the desire has some causal genesis. What is denied is that there is a “reasons explanation” for the desire: the subject can’t (truly) give a (non-circular) reason for wanting the thing.

And this despite the fact that the sensation is interesting to you in virtue of its intrinsic features.

Sumner was evidently misled: from the fact that you desired the pins and needles sensation for its intrinsic qualities, Sumner inferred – fallaciously, I say – that your desire for the pins and needles sensation was intrinsic.

That is, the following is a fallacy:

You desire something for its intrinsic qualities.  
∴ Your desire is intrinsic.

To see that this is a fallacy: suppose you want to own some copy or other of *The Methods of Ethics*. Suppose the book in your hand is a copy of *The Methods of Ethics*. So you want it. That is, you want that you own the book in your hand. Now, in virtue of what is the book in your hand a copy of *The Methods of Ethics*? It is a copy of *The Methods of Ethics* in virtue of its intrinsic features (we can suppose). So, you want the book in your hand for its intrinsic qualities.<sup>24</sup> Nevertheless, your desire for the book in your hand is extrinsic. You want the book in your hand only because you want to own a copy of *The Methods of Ethics* and the book in your hand is a copy of *The Methods of Ethics*.

---

<sup>24</sup> In order to make this point, I have been speaking as if concrete objects, like books, are the objects of desire. I do this because the point involves the notion (employed by Sumner in his objection) of desiring an object for some intrinsic feature of the object. And this notion evidently presupposes that the objects of desire can be concrete objects. Those like me who assume that states of affairs are the objects of desire can make sense of Sumner's notion by defining a derivative sense of 'desire'. When a person S desires the proposition that S has an F or that S has b, we can say that S *derivatively desires* an F or that S *derivatively desires* b.



#### 5.6.4 “I want it because it’s pleasant,” or: A Euthyphro Problem

In responding to the interesting sensation objection, I have employed a test for determining whether a desire is intrinsic: if the subject cannot give a reason for her desire, or if the reason is in some way circular, then the desire is probably intrinsic; if the subject has no trouble giving a non-circular reason for her desire, then the desire is probably extrinsic. But consider what might happen when a subject is asked to explain her desire for some pleasant sensation. The subject might say she wants to experience the sensation *because it is pleasant*. I appealed to an analogous fact above in the interesting sensation case to establish that the desire for the sensation in that case was extrinsic. Perhaps foes of reduction can appeal to the corresponding fact in the pleasant sensation case to establish that the desire in that case is likewise extrinsic. If so, then the opponent may have given us reason to think that we *never* have intrinsic desires for sensations (or perhaps for anything!, for desires for sensations are the paradigmatic case of intrinsic desires). Perhaps our chains of extrinsic desire always end not in intrinsic desire but in pleasure. This line of thought is incompatible with a reduction of pleasure to desire.

I think we can get at the same objection in a different way, by asking a Socratic question. Do we desire pleasant sensations because they are pleasant, or are pleasant sensations pleasant because we desire them? Although I think our confidence in it is less than is our confidence in the intuitive answer to Socrates’s actual question (about piety, or moral rightness), the more intuitive answer to this question is probably that we desire pleasant sensations because they are pleasant. But this conflicts with the

reduction of sensory pleasure to desire, according to which pleasant sensations are pleasant because they are desired.<sup>25</sup>

I respond that although the motivational theory does take the less intuitive horn of the Socratic dilemma, the theory is still compatible with an important sense in which we desire pleasant sensations because they are pleasant. In ordinary circumstances, when someone says, “I want to taste that beer because I find the taste pleasant,” they say something compatible with the motivational theory of sensory pleasure. Or at least there is a natural interpretation of this remark that is compatible with the theory.

When a person says, “I want to taste that beer because I find the taste pleasant,” he may mean roughly the following, “I want to taste that beer because when I taste that beer I will find the taste I get pleasant.” But this is no problem for the reduction of sensory pleasure to desire. According to the theory, when a person says this, he says, “I want to taste that beer because when I taste that beer I will be intrinsically desiring the taste I get.” Of course, the person might not realize or even believe he is saying this, but this is just because he doesn’t accept or perhaps even know about the motivational theory of sensory pleasure. If Frege is right, then statements about numbers are statements about sets. It is no argument against the reduction of numbers to sets that people who speak about numbers don’t realize they are talking about sets. Likewise, it is no argument against the reduction of pleasure to desire that people who speak about pleasure don’t realize they are talking about desire.

---

<sup>25</sup> Rachels (2000, p. 192) makes more or less this objection: “People want their unpleasure to end because it’s unpleasant. However, on [a desire theory of pleasure and pain], this amounts to saying that people want their unpleasure to end because they want it to end, which is nonsense.”

According to this reply, our Socrates can be interpreted as presenting a false dilemma. He asks, “Do we desire pleasant sensations because they are pleasant, or are pleasant sensations pleasant because we desire them?” The reducer of sensory pleasure to desire responds, “Yes and Yes.” Yes, we desire pleasant sensations because they are pleasant (in other words, we desire them in advance because we know we will desire them when we get them). And Yes, pleasant sensation qualify as pleasant in virtue of the fact that they are intrinsically desired.

This reply reveals that the Socratic question is ambiguous. Presumably what our Socrates really meant is the following: “When a person is experiencing and simultaneously desiring to be experiencing some pleasant sensation, does he desire to be experiencing the sensation because it is independently pleasant, or does the sensation qualify as pleasant only in virtue of the fact that the subject is simultaneously desiring it.” To this clunky question, the reducer of pleasure to desire must indeed take the latter, less intuitive horn.

But this should not be terribly surprising, for that is a common price of reduction. To assess how costly the less intuitive horn is, we weigh its counter-intuitiveness against the explanatory and theoretical benefits of reduction. In the moral case, I side with Socrates. I find it too central to our concept of morality that an action cannot be made morally right simply in virtue of its being permitted by some authority – even God. Moreover, there are other explanatory unifications of right action – normative theories of rightness – available.

Things are different in the sensory pleasure case. First of all, the anti-Euthyphronic intuition in the sensory pleasure case is less compelling than the anti-

Euthyphronic intuition in the moral case. To see this, consider to what the real Euthyphro commits himself in claiming that right acts are right in virtue of being approved by God. Euthyphro must accept counterfactuals like the following:

**CF1:** If God had approved of Ted Bundy's rape and murder of Joni Lenz, then Ted Bundy's rape and murder of Joni Lenz would have been morally permissible.<sup>26</sup>

CF1 is hard to swallow. By contrast, reducers of sensory pleasure to desire need only accept counterfactuals like the following:

**CF2:** If Jack didn't want (intrinsically) to be feeling the sensations caused by that massage, then those sensations would not have been sensations of pleasure.

**CF3:** If Zoe didn't want (intrinsically) to be tasting the taste of that peach, then the taste Zoe experienced would not have been pleasant.

CF2 and CF3 are nothing like CF1. CF1 is incredible; intuition is much less clear about CF2 and CF3.

Secondly, I see no plausible alternative explanatory unifications in the offing. If we hold our anti-Euthyphronic intuition sacrosanct in the sensory pleasure case, we seem to be forced into the unwelcome position of regarding sensory pleasantness to be an inexplicable primitive property (as Felt-Quality Theories have it) or an evaluative property (as Desirability Theories have it). Better to stick with theory on this point, and

---

<sup>26</sup> In assessing this counterfactual (or perhaps it's a counterpossible, but that's ok), we must take care not to interpret it as a *back tracker* (or something analogous). We cannot justify it, for example, with, "If God had approved of Ted Bundy's rape and battery of Joni Lenz, then it must be that that act was part of some worthy divine plan that we cannot hope to understand." The relevant counterfactual requires that we hold everything constant – including the actual fact that the act was not part of some divine plan and has no good consequences – changing only God's evaluation of the act.

learn from it that a common idea – that we desire pleasant sensations because they are pleasant – is mistaken.

### 5.6.5 An Argument from Undesired Non-Pains

RSPD3 is a theory about pleasure only. But I have suggested above that I would endorse a reduction of pain to desire as well. Even if a reduction of sensory pleasure to desire is successful, we are still left with a heterogeneity problem for sensory pain, and we still need to explain the way in which sensory pain is the opposite of pleasure. A motivational theory of sensory pain promises to solve both problems.

We formulate this reduction of sensory pain to desire as follows:

**RSPD3’:** a sensation *S*’s instantiating *P* at *t* is a sensory pain =df.  
S intrinsically desires at *t* that S does not instantiate *P* at *t*.

On this view, a sensation gets to be a sensation of pain just in case its subject wants intrinsically *not* to be feeling it while she is feeling it.

There is a special difficulty for RSPD3’, one for which there is apparently no analogue in the pleasure case. It seems there are sensations that are *uncomfortable*, even *unpleasant*, without being *painful*. The trouble this makes for a motivational theory of sensory pain should be obvious. Any uncomfortable or unpleasant sensation we would surely intrinsically desire not to be feeling. RSPD3’ implies therefore that such sensations are painful. But not all such sensations are painful. We are more restrictive in our ascriptions of sensory pain. We are reluctant to call many undesirable sensations ‘painful’. We may say that they are uncomfortable, or that we don’t like them, or that they are unpleasant, while steadfastly denying they are *painful*.



Consider a mild itch. We don't always like the way these feel. They can be uncomfortable and annoying. We want not to be feeling them – that's why we scratch them. But are these mild itches *painful*? Are mild itch sensations sensations of pain? No – it seems like an abuse of language to call a mild itch 'painful'. The word is just not used this way. Toothaches are properly called 'painful'; paper cuts are painful; having my wound cleaned with rubbing alcohol was painful. But not mild itches; they're merely unpleasant, uncomfortable.<sup>27</sup>

That there are uncomfortable and unpleasant sensations that are not painful is supported by that common question of doctors: "Do you feel any pain or discomfort in your \_\_\_\_\_?" It doesn't seem that the phrase 'pain or discomfort' is redundant. Sumner (1998, pp. 177-78) has recognized this phenomenon as well, and provides some nice examples:

... there are many feelings (dizziness, nausea, disorientation, apnea, etc.) which we are intrinsically displeased to be having but which we would not commonly describe as painful.

Sidgwick also seemed to recognize the possibility of desiring not to be experiencing some feeling without the feeling's being painful. He considers (1907, p. 46), "The question whether all [unsatisfied] desire has in some degree the quality of pain" and has

no hesitation in answering the question in the negative. Consider again the case of hunger; I certainly do not find hunger as an element of my normal life at all a painful feeling: it only becomes

---

<sup>27</sup> One might suppose that painfulness is just intense enough unpleasantness, but this seems mistaken. There are very mild sensory pains – such as a pin-prick to the forearm – and very intense sensory discomforts – such as intense queasiness – that I would hesitate to describe as sensory pains.

painful when I am in ill health, or when the satisfaction of the appetite is abnormally delayed.<sup>28</sup>

I think the argument under consideration successfully refutes RSPD3'. It's just not the case that being intrinsically averse to some sensation guarantees that the sensation is painful.

My response to refutation: change the subject. Revise the subject matter of the reduction of sensory pain to desire. Change the left-hand side of the analysis rather than the right-hand side. Make it a theory about sensory "unpleasantness" rather than about sensory pain. According to the new theory, a sensation gets to be an *unpleasant sensation* when its subject intrinsically desires not to be feeling it while he is feeling it:

**RSUD:** a sensation *S's instantiating P at t* is unpleasant =df.  
S intrinsically desires at t that S does not instantiate P at t.

Even if the present objection does refute the motivational theory of *sensory pain*, it fails against the motivational theory of *sensory unpleasantness*. It does not seem possible for there to be sensations that we intrinsically desire not to be feeling that are nevertheless not unpleasant. If we intrinsically want not to be feeling some sensation, that sensation, I submit, is unpleasant. The sensation of a mild itch, or of dizziness, or of nausea – these might not be painful, but each is definitely unpleasant.

I can anticipate three complaints about changing the subject in this way. The first complaint is that the new theory is true by mere stipulation and therefore uninteresting. We began with a quest to understand sensory pain. That proved difficult.

---

<sup>28</sup> Sprigge (1988, pp. 128-129) and Rachels (2004, p. 248) also recognize non-painful unpleasantness. They prefer 'unpleasure' to 'unpleasantness'. More on this below.

So we coined a new term – ‘sensory unpleasantness’ – that makes the motivational theory automatically true.

I think this first complaint is pretty clearly misguided. The concept of a sensation’s being unpleasant was not cooked up in the philosophy lab. The term ‘unpleasant’ is a term of ordinary language, the corresponding concept is one of ordinary thought, and it applies to sensations in a familiar way. The motivational theory of this ordinary concept of sensory unpleasantness, if it is true, is certainly not true by stipulation. We can even imagine how it could be proven false: it needs to be shown that there are sensations we would clearly classify as unpleasant but to which their sensor is clearly not intrinsically averse.

The second complaint involves the implications for the oppositeness problem.<sup>29</sup> I suggested it is a datum that pleasure is the opposite of pain. By changing our focus, am I not implying that pain is in fact not the opposite of pleasure, and thereby denying a undeniable datum?

I *am* implying that pain is in fact not the opposite of pleasure, but I am not thereby denying a undeniable datum, because the very observation that motivates our change of focus also proves it is just false that pain is the opposite of pleasure. It shows that the opposite of pleasure must be something more inclusive. And notice that changing the subject in this way doesn’t threaten the desire-theoretic solution to the oppositeness problem. For that solution to stand, all that we need is for the opposite of

---

<sup>29</sup> I owe this objection to Daniel Howard-Snyder.

sensory pleasure to be susceptible to a reduction to desire analogous to the reduction of sensory pleasure to desire. I claim this is exactly what we have.<sup>30</sup>

The third complaint about changing the subject is the best one. It is that we are no longer talking about the right concept, the concept that plays the important role in moral philosophy, the concept that has moral, prudential, and rational significance. It is the concept of *pain* that we wanted to understand, and we wanted to understand it, the objection goes, because that's the concept that figures in hedonistic theories of value and welfare, in utilitarian theories of obligation, in pluralistic theories of value, welfare, and obligation that contain hedonistic or utilitarian components, and in midlevel moral principles making use of the notion of inflicting pain. If I change the subject from pain to unpleasantness, haven't I failed in my promise to shed light on a central concept of moral philosophy?

No. Ironically, if we admit the possibility of unpleasant non-pains, then consideration of this "conceptual role" argument reveals something unintended. It reveals that painfulness isn't quite the concept that has moral, prudential, and rational significance after all. It is unpleasantness. Sensory unpleasantness is far better suited to play the role in moral philosophy we thought would be played by sensory pain.

---

<sup>30</sup> Stuart Rachels (2004, pp. 247-48) claims that *unpleasure* – not unpleasantness – is the opposite of pleasure. One reason to favor Rachels's view is that perhaps there are painful sensations – and so sensations that belong in the opposite-of-pleasure category – that don't count as unpleasant, because very intense pains are not properly called 'unpleasant'. But given our purposes here, I am hesitant to use the expression 'sensory unpleasure'. The word 'unpleasure' is so obscure that a theory about the property it expresses may be difficult to evaluate. Although 'unpleasure' is no philosophical neologism (interestingly, the OED's first definition of it is *unpleasantness!*), if we use it to change the subject, we may expose ourselves to the first complaint above. Moreover, I'm not convinced it is literally false to call extremely intense pains 'unpleasant'.

To see why, consider the following simple moral principle:

**Non-Maleficence:** It is morally wrong to inflict needless pain on innocent people.

Forget for a moment our present nick-pickings about the concept of sensory pain and get in tune with the spirit of Non-Maleficence. Imagine the sorts of behavior the principle is supposed to prohibit. Now consider the case of Tina the careful torturer. Tina abducts innocent victims and subjects them to her bizarre torture. She makes her victims wear itchy wool sweaters with nothing underneath and won't let them scratch. She feeds her victims rotten food, causing them terrible nausea. She dunks their heads into the tub and won't let them come up for air, causing unbearable feelings of apnea. Now recall the spirit of the simple principle prohibiting needless pain. Is the principle supposed to condemn Tina's behavior? Of course it is.

But recall the puzzle at hand about sensory pain. Recall that itches can be very unpleasant without technically being painful. Likewise for nausea and apnea. Put this to bear on the case of Tina the careful torturer. Tina (perhaps accepting the principle above and knowing about the present debate) is very careful never to cause what would qualify as sensory pain. This is why we call her torture 'careful'. What follows from this, it seems to me, is not that Tina's behavior escapes reproach on a technicality. What we see, rather, is that the simple principle was meant to prohibit Tina's behavior. If this doesn't show that the meaning of the word 'pain' is in need of explication, it



shows that proponents of Non-Maleficence really have unpleasantness and not pain in mind.<sup>31</sup>

The same line of thought applies to more theoretical positions as well. Obviously, when utilitarians tell us to maximize the balance of pleasure over pain, the intended notion of pain is broad enough to include Tina's torture victims in the hedonic calculus. Obviously, when hedonists say that pain is bad, the intended notion of pain is broad enough for hedonists to recognize that things go badly for Tina's victims.

And this broader notion of pain – one that includes unpleasantness – has this feature, crucial for our larger purpose: the motivational theory about *that* concept is plausible. The concept of pain that features in our common sense moral thought as well as in our more sophisticated moral thought is a concept of pain that bears a conceptual connection to desire. This, I take it, was exactly Parfit's point when he said (1984, p. 493):

On the use of 'pain' which has rational and moral significance, all pains are when experienced unwanted ... .

Parfit's view might be that 'pain' has no settled meaning (and so it is unsettled whether there really are unpleasant non-pains), but that one legitimate way to settle its meaning gives the corresponding concept rational and moral significance.

So not only is it permissible to change the subject from sensory pain to sensory unpleasantness, it is required. Sensory unpleasantness, not sensory pain, is what we must understand if we wish to understand the concept that plays the important role in moral philosophy.

---

<sup>31</sup> Sprigge (1988, pp. 128-29) is aware of this problem with the word 'pain'; he opts to keep the word and stipulate a broader meaning for it.

I do grant that some mystery remains. We have the class of unpleasant sensations. A subset of the unpleasant sensations qualify as pain sensations. Why? What is it about the pain sensations in virtue of which they are properly described as 'painful'. What distinguishes an unpleasant sensation that is also painful from one that is not? Why has our language and thought delineated this subclass of unpleasant sensation and given it a name, especially given that the distinction has no moral or rational import? I do not know the answer to these questions.

## CHAPTER 6

### DESIRE SATISFACTIONISM AND HEDONISM

Hedonism and the desire theory of welfare (or “desire satisfactionism,” as I will call it) are typically seen as archrivals in the contest over identifying what makes one’s life go best. Hedonism identifies the good life with the pleasurable life. Desire satisfactionism is supposed to be less paternalistic: you don’t need to get pleasure to be well off; you just need to get what you want, whatever it is. It is surprising, then, that the most plausible form of hedonism is extensionally equivalent to the most plausible form of desire satisfactionism. It is even more surprising that the most plausible form of hedonism *just is* the most plausible form of desire satisfactionism. I intend here to defend this surprising thesis. How can a single theory of welfare be a version of both hedonism and desire satisfactionism? The answer has to do with pleasure. But it’s no dubious psychological claim about how we humans ultimately desire only pleasure. The answer, rather, lies in what pleasure *is*: pleasure, I will argue, is the subjective satisfaction of desire.

So in this paper I am trying to accomplish three major tasks: (i) discover the most plausible form of the desire theory of welfare, (ii) discover the most plausible form of hedonism, and (iii) defend a theory about the nature of pleasure. Since each of these topics deserves at least its own paper, my treatments of them will be overbrief. I will not be able to talk about every argument against each of the three main views under discussion. And the objections I do discuss I must discuss quickly. Nevertheless, I

hope to make a decent preliminary case for the thesis that these two rival approaches to the nature of welfare come to the same thing.

## 6.1 Desire Satisfactionism

If hedonism tells us that enjoying what we get makes our lives go well, desire satisfactionism tells us that getting what we want makes our lives go well. We might prefer a desire theory of welfare when we consider the fact that we value many things other than pleasure, such as friendship, love, truth, beauty, freedom, privacy, achievement, solitude—the list is long. If one's life is filled with such things, and they are exactly the things one wants, and one therefore thinks that one's life is pretty good, it seems arrogant and paternalistic of the hedonist to insist otherwise, just because the life lacks enough pleasurable experience.

The simplest form of the desire theory of welfare I call *Simple Desire Satisfactionism*. According to Simple Desire Satisfactionism, your life goes well to the extent that your desires are satisfied. Every time a subject S desires that some state of affairs p be the case, and p is the case, S's desire that p be the case has been satisfied. And according to Simple Desire Satisfactionism, he is thereby better off. It is no part of Simple Desire Satisfactionism that, for a person's desire to be satisfied, the person must experience feelings of satisfaction.

Everyone seems to agree on one restriction to Simple Desire Satisfactionism right off the bat: we should count only intrinsic desires. If I want to turn on my CD player only because I want to hear the Pixies, I'm made no better off if only the first desire is satisfied. That desire is extrinsic: I have it only because I desire something

else, and I think the thing extrinsically desired will lead to that something else.

Something goes well for me in this situation only if my intrinsic desire to hear the Pixies is satisfied. Let *Intrinsic Desire Satisfactionism* be the desire theory of welfare that counts only the satisfaction of intrinsic desire.

Now on to more interesting problems for the desire approach to welfare. I begin with two objections that I think are successful but that point to improved versions of the theory. I conclude this section with two popular objections that I think are unsuccessful.

### 6.1.1 The Problem of Changing Desires

Richard Brandt, a one-time (1979, ch. 6) defender of desire satisfactionism, became convinced later (1982, p. 179) that any form of desire satisfactionism suffers from an irremediable defect: there is no satisfactory way to handle cases in which a desire for something is unstable. Suppose for my whole life I want rock and roll on my 50<sup>th</sup> birthday; suppose a week before the birthday my tastes change and I want easy listening on my birthday (and will continue to want easy listening). Intrinsic Desire Satisfactionism seems to imply that we make me better off by giving me rock and roll on my 50<sup>th</sup> birthday. But this seems mistaken – the theory suggests we force-feed people things they no longer want. But we can't focus exclusively on present desires either (as has been suggested). Suppose I want pushpin now but my desire for pushpin will last only another day, and then for the rest of my life I will want poetry. You don't make me better off by permanently giving me pushpin instead of permanently giving me poetry.



Some philosophers try to handle this problem by moving to *ideal desires* – that is, by counting only the satisfactions of the desires we *would* have if we were better informed, or were more rational. Indeed, I think it is fair to say that ideal desires are the *deus ex machina* of the desire satisfaction literature, dragged in at any moment to solve whatever objection is forthcoming. It is hoped that one's ideal desires are unchanging, and so The Problem of Changing Desires would not arise. But I don't think such a theory will get desires stable enough. Certainly there is nothing incoherent in the idea of a fully informed person changing his intrinsic preferences.

Brandt abandoned the desire approach wholesale, but a desire theoretic solution can be had: say that a state of affairs is a desire satisfaction only if it is a case of a person wanting something and getting it *at the same time*. Call the resulting theory *Concurrent Intrinsic Desire Satisfactionism*. Such a theory generates the correct results in the two cases above. Since it ignores past preferences, it implies in the first case that we provide the most benefit by giving me easy listening. In the second case, the theory implies that we benefit me more by giving me poetry since, as the years go by, I will get what I am wanting at each time. This far outweighs the fact that I don't get what I want just for today.

I admit that at this stage the concept of concurrence may be less than perfectly clear. For instance, can present desires about the future or past (so-called now-for-then desires) ever be concurrently satisfied? Since Concurrent Intrinsic Desire Satisfactionism isn't the theory I like best anyway, I'll forgo clarifying the doctrine.

Our intuitive and imperfect understanding of concurrence should be enough for the time being.<sup>1</sup>

### 6.1.2 The Objection from Remote Desires

The name I give this objection derives from the remoteness of the *objects* of desire. Sometimes we desire certain remote (in time, place, importance, etc.) states of affairs to obtain. James Griffin presents this kind of objection in the following passage about desire satisfactionism (1986, pp. 16-17):

The breadth of the account, which is its attraction, is also its greatest flaw. ... It allows my utility to be determined by things that I am not aware of (that seems right: if you cheat me out of an inheritance that I never expected, I might not know but still be worse off for it), but also by things that do not affect my life in any way at all. The trouble is that one's desires spread themselves so widely over the world that their objects extend far outside the bound of what, with any plausibility, one could take as touching one's well-being.

Shelly Kagan (1998, p. 37) presents a similar line of thought:

... according to the preference theory, I am made better off by the satisfaction of my various desires, regardless of the subject matter of the given desire. Suppose, then, that I am a large fan of prime numbers, and so I hope and desire that the total number of atoms in the universe is prime. Imagine, furthermore, that the total number of atoms in the universe is, in point of fact, prime. Since this desire is satisfied, the preference theory must say that I am better off for it ... . But this is absurd! The number of atoms in the universe has nothing at all to do with the quality of my life. ... So the preference theory must be false.

And Derek Parfit has made this objection in the form of the case of The Stranger on the Train (1984, p. 494):<sup>2</sup>

---

<sup>1</sup> I am indebted to Dick Arneson here.

Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured. On the Unrestricted Desire-Fulfillment Theory, this event is good for me, and makes my life go better. This is not plausible. We should reject this theory.

The solution Parfit offers the desire satisfactionist is to restrict the theory to count only desires “about our own lives.” Parfit admits, however (1984, p. 494), that “when this theory appeals only to desires that are about our own lives, it may be unclear what this excludes.” My own view is that even if it can be made clear what it means for a desire to be about one’s own life – and Mark Carl Overvold (1980, 10n) has offered a way – we are barking up the wrong tree, because many desires not about our own lives are important, too. So I propose an alternative solution to The Objection from Remote Desires: count a desire satisfaction only if the subject is aware that the desire is satisfied. The theory based on this restriction will be explained shortly.<sup>3</sup>

### 6.1.3 The Problem of Desiring Not To Be Well-Off

Imagine a man who, ridden with guilt for past crimes, wants (intrinsically) to be badly off. In order to satisfy this desire, the man takes an arduous, boring, and

---

<sup>2</sup> Actually, there is no train in Parfit’s original example, but the literature has since supplied one.

<sup>3</sup> James Griffin considers another way to restrict desire satisfactionism so as to handle The Objection from Remote Desires (1986, p. 17-23, *passim*). He suggests that we count a desire only if it is “one of [our] aims,” “one of [our] central ends” (p. 21). Perhaps these notions can be cashed out in terms of the *intensity* of the desire. Griffin here is assuming that the force of the objection comes from the remoteness of the importance of the desire – they are remote to what we most care about.

insignificant job.<sup>4</sup> He's pretty miserable. He seems to have succeeded in getting what he wants: being badly off. But since he is getting what he intrinsically wants while he wants it, Concurrent Intrinsic Desire Satisfactionism seems to imply, absurdly, that he is well-off.

I don't think The Problem of Desiring Not To Be Well-Off works. I deny that Concurrent Intrinsic Desire Satisfactionism implies that the man described above is well-off. My reply depends in part upon some views about the nature of pleasure/happiness and pain/misery that will emerge later in the paper. The short answer is that it is impossible (conceptually, metaphysically) to experience things like misery, boredom, arduousness, etc. without having desires frustrated.

The example may help to clarify. Here is how the actual desire-satisfaction theorist should describe what goes on in the case. The man has a job that is arduous, boring, and insignificant. His daily life is thus jam-packed with desire frustration: to feel bored is, necessarily, to want (intrinsically) to be doing something other than what one is doing. For a person to find a task to be arduous is also, necessarily, for him to have certain desires frustrated. Does it make sense to say that someone found some stretch of time boring, or arduous, but that he was getting everything he wanted at every moment of that stretch of time?

So each day at this job, the man is miserable. His life is filled with desire frustration. But that's not the whole story, for he does have at least one desire satisfaction on his plate: the desire that he be badly off. That's one point in his favor

---

<sup>4</sup> This example is Richard Kraut's (1994). See also Robert Merrihew Adams (1999, pp. 83-93).

against all the points against him. But the satisfaction of this desire to be badly off must, of necessity, count for less, in terms of welfare, than all the daily frustrations he racks up. If it were otherwise, then the man wouldn't be badly off, and the desire to be badly off would no longer be satisfied.<sup>5</sup> So Concurrent Intrinsic Desire Satisfactionism implies, correctly, that the man is not well off, that he has succeeded in becoming badly off.

#### 6.1.4 The Problem of Defective Desires

The Problem of Defective Desires attempts to shake desire satisfactionism at its foundations. The problem is that it seems that we can desire things that are bad for us – things such that, if we get them, we are made worse off because of it. Call a desire 'defective' if it would not be good for the subject to satisfy it. The claim is that there are indeed defective desires, but that, according to an unrestricted theory like Concurrent Intrinsic Desire Satisfactionism, there are none because, on the theory, all desire satisfactions are good.

It is not hard to think of cases in which satisfying a desire is bad for us. We have, for example, *ill-informed desires*: we might have a desire to drink from the river not knowing that it will make us sick. The thought is that desire satisfactionism

---

<sup>5</sup> Which, lo and behold, could tilt the balance back again and make him badly off. But then the desire to be badly off *would* be satisfied, and the man would then fail once again to be badly off. But then once again the desire *wouldn't* be satisfied. In short, we have a paradox. It is beyond my scope here to attempt to solve this paradox. But I will say that there is a paradox in the vicinity that is *everyone's* paradox (whether he is a desire satisfactionist or not). It arises in cases in which a subject desires that his desires not be satisfied.



implies, absurdly, that it would be good for us to take the drink.<sup>6</sup> There are also *irrational desires*: a person might know that going to the dentist is in his interest, but still he prefers and chooses not to go, because he is weak-willed. The claim is that desire satisfactionism implies, incorrectly, that since he desires not to go, and all desire satisfactions are good for a person, it is good for him not to go to the dentist.<sup>7</sup>

The problem is brought out by the fact that it is very natural to evaluate and criticize desires and desire satisfactions themselves. But desire satisfactionism, according to which all desire satisfactions are good for the subject, doesn't seem to allow it.

Some desire satisfactionists accept the argument from defective desires, and, in response, develop a version of ideal desire satisfactionism. Richard Brandt's (1979) version of the theory counts not our actual desires, but of our "rational desires" – roughly, the desires we would have if we were fully informed. In my view, ideal desire satisfactionism is both ineffectual in solving The Problem of Defective Desires and fails in its own right. It is ineffectual in solving The Problem of Defective Desires because there is no guarantee that idealization will remove all the offensive desires (e.g., irrational desires (discussed above) and base desires (to be discussed below) can survive full information). The theory fails in its own right because (i) unwanted satisfactions of merely ideal desires are not, contrary to what the theory says, necessarily intrinsically

---

<sup>6</sup> That is, that the *satisfaction* of the desire for drink would be good. In my opinion, properly formulated desire theories say that desire *satisfactions*, not the *objects* of desire, are intrinsically good for their subject. This will become clearer later, when the official formulations of the final theories are given. The example is from Carson (2000, pp. 72-73).

<sup>7</sup> Schwartz (1982, p. 196) makes more or less this argument.

good for a person<sup>8</sup>, and (ii) there are a host of problems surrounding the concept and process of idealization.<sup>9</sup> One could devote at least a paper to the development and defense of these claims; I must leave them undeveloped and undefended here.

In my view, the Problem of Defective Desires is soluble fairly straightforwardly. Standard “actualist” versions of desire satisfactionism *are* compatible with the criticism of desires, though all criticism must be “extrinsic.” Extrinsic desires can be criticized for failing to be efficient means to the satisfaction of intrinsic desires. Intrinsic desires can be criticized for leading to a less favorable balance of satisfactions over frustrations than would have occurred had the intrinsic desire not been had, or not been satisfied. In both cases, criticism is “extrinsic”: the desires, or their satisfactions, are criticized not in themselves, but for what they lead to, or for what they fail to lead to.

We can make this more clear and explicit by distinguishing *intrinsic goodness* (for a subject) from *all things considered goodness* (for a subject). A state of affairs *p* is *intrinsically good for a subject S* iff it is good in itself for *S*; iff given two lives exactly the same except with respect to *p*, the *p*-life is a better life for the person (i.e., contains more welfare) than the not-*p*-life. Standard actualist forms of desire satisfactionism

---

<sup>8</sup> This complaint is made in many places. See, e.g., Griffin (1986, pp. 11-12), Sobel (1994, pp. 792-793), Tännsjö (1988, pp. 87-88) and Feldman (2004, p. 17). Railton (1986, p. 16) and Carson (2000, p. 226) attempt to solve the problem by counting not what one’s ideal self wants but rather what one’s ideal self wants one’s actual self to want. I find this reply unconvincing because nothing guarantees that one’s ideal self would be benevolently disposed towards one’s actual self – he might be indifferent (or worse) to his counterpart’s welfare.

<sup>9</sup> Such problems have been discussed in quite a few places. See Adams (1999, 86-87), Arneson (1999, pp. 127-130, 133-135), Carson (2000, p. 226-230), Gibbard (1990, pp. 18-22), Loeb (1995, pp 19-20), Rosati (1995, pp. 307-324), Sobel (1994, pp. 794-807), and Velleman (1988, 365-371).

(like Concurrent Intrinsic Desire Satisfactionism) entail that all (concurrent, intrinsic) desire satisfactions are intrinsically good for their subject.

A state of affairs *p* is *all things considered good for a subject S* iff the life *S* would lead were *p* to obtain is better than the life *S* would lead were *p* not to obtain. Since it can and does happen that a state of affairs that is intrinsically good for a subject leads to states of affairs that are intrinsically bad for the subject, *being intrinsically good* does not entail *being all things considered good*. Thus, standard actualist forms of desire satisfactionism (like Concurrent Intrinsic Desire Satisfactionism) do *not* entail that all (concurrent, intrinsic) desire satisfactions are all things considered good for their subject.

Armed with this distinction, we can see how an actualist desire satisfaction theory can accommodate the obvious fact that we can desire things that are bad for us. Recall the case of the polluted river. The desire satisfactionist should say that it was intrinsically good for the subject to quench his thirst by drinking from the polluted river, since an intrinsic desire (the thirst) was satisfied. But she can and should also say that that desire was defective and should not have been satisfied since it was all things considered bad to drink from that river.<sup>10</sup> It was not in the subject's long-term self interest. And the desire satisfactionist can say this. She can say it because drinking from the river led to so many desire frustrations – those accompanying being sick.

A similar response is available with respect to irrational desire. We all agree, including even the irrational subject, that it would be good for him to go to the dentist,

---

<sup>10</sup> Let me emphasize that to call a desire 'defective' in this context is not to say that it is therefore appropriate to condemn the desirer for having it. It might have been a perfectly reasonable desire to satisfy given the desirer's evidence.

and that he should go. Even though doing so will frustrate one of the subject's intrinsic desires, actualist desire satisfaction theories still imply it would be good for him to go. It would be all things considered good, since the life the subject would lead were he to go to the dentist is better for him than the life he would lead were he not to go.

So critics of the desire theory of welfare do present cases of defective desires, but none so far is a convincing case in which an intrinsic desire is intrinsically bad to satisfy.

But there another kind of case that is clearly meant to be an intrinsic criticism of an intrinsic desire: the case of *base desires*. G.E. Moore's (1993, pp. 146-147) famous example of base pleasures can be modified for our purposes: imagine a case in which a person's strongest desires would be satisfied by a perpetual indulgence in bestiality. So long as the imagined indulgence doesn't bring with it concomitant frustrations, a theory like Concurrent Intrinsic Desire Satisfactionism must deem the life a very good one. Since an analogous objection faces the hedonist, I save discussion of this objection for the upcoming section on hedonism.

For now, I conclude that at least the first two variants of the Problem of Defective Desires do not refute Concurrent Intrinsic Desire Satisfactionism.

## 6.2 Subjective Desire Satisfactionism

This dialectic leads us to a theory I will call *Subjective Desire Satisfactionism*. More standard versions of desire satisfactionism say, in a nutshell, that welfare consists

in getting what one wants. Subjective Desire Satisfactionism says, on the other hand, that welfare consists in believing one is getting what one wants.<sup>11</sup>

We can state the theory better as follows. An instance of “subjective desire satisfaction” is a state of affairs in which a subject (i) has an intrinsic desire at some time for some state of affairs and (ii) believes at that time that the state of affairs obtains. An instance of “subjective desire frustration” occurs when (i) above holds but the subject believes that the desired state of affairs does *not* obtain. The value for the subject of (or the amount of welfare in) a subjective desire satisfaction is equal to the intensity of the desire satisfied. Likewise for frustrations, except that the number is negative.<sup>12</sup> The theory is summative so that the total amount of welfare in a life is equal to the sum of the values of all the subjective desire satisfactions and frustrations in that life.

Officially, here is the theory:

**Subjective Desire Satisfactionism:**

- (i) Every instance of subjective desire satisfaction is intrinsically good for its subject.
- (ii) Every instance of subjective desire frustration is intrinsically bad for its subject.

---

<sup>11</sup> Earlier, I glossed the theory using the stronger concept of *being aware that* (rather than merely *believing that*, which I use here). The first and simplest formulation of the theory will use belief; it will require only that the subject believe that what he wants obtains (and will ignore the issue of whether it actually obtains). In the upcoming section on hedonism, I will consider adding a truth requirement to both subjective desire satisfactionism and the hedonist theory.

<sup>12</sup> One modification worthy of exploration includes the recognition of intensities of belief as well. I have in mind the common idea that belief is not an all or nothing affair – that there are degrees of belief. The modified theory would hold that the value of a subjective desire satisfaction is a function of the intensity of the belief and the intensity of the desire.



- (iii) The intrinsic value for its subject of an instance of subjective desire satisfaction = the intensity of the desire satisfied.
- (iv) The intrinsic value for its subject of an instance of subjective desire frustration =  $-(\text{the intensity of the desire frustrated})$ .
- (v) The intrinsic value of a life for the one who lives it = the sum of the values of all the instances of subjective desire satisfaction and frustration contained therein.

The phrase ‘the intrinsic value of a life for the one who lives it’ means the same as ‘the total amount of welfare in a life’ or ‘how good the life was for the one who lived it’.

Subjective Desire Satisfactionism, or SDS for short, avoids The Problem of Changing Desires in a manner similar to the way in which Concurrent Intrinsic Desire Satisfactionism did. According to SDS, a person’s past desires have no direct impact on the person’s welfare – we do not benefit a person simply in virtue of satisfying (subjectively or otherwise) a formerly held desire. (Incidentally, that a person’s past desires have no direct impact on welfare does not mean that we should ignore a person’s past desires. For they are often decent indicators of the desires she has now or will have in the future.) In the case discussed earlier, SDS implies that it would be best to give me easy listening on my 50<sup>th</sup> birthday. That is what I will want on the day, and actually giving it to me is the best way to ensure that I will think I am getting it. It was left unclear what Concurrent Intrinsic Desire Satisfactionism had to say about desires about the future and past (the now-for-then desires). On SDS the matter is clear. It is good for a subject to subjectively satisfy her desires about any time, past, present, and future. Now-for-then desires count. This has an interesting, and I think quite plausible, consequence for the case about the 50<sup>th</sup> birthday party. Throughout my life, I want that I get rock and roll on my 50<sup>th</sup> birthday. SDS implies that it is good for me to believe, throughout my life, that that’s what I will get. As the day nears and my preference

changes to easy listening, SDS implies that it is good for me to believe that easy listening is what I will get. Optimistic anticipation is a good thing.

According to SDS, a person racks up welfare only by subjectively satisfying his present desires; it follows that posthumous satisfactions are worthless, at least with respect to their welfare value. But this is not because SDS ignores now-for-then desires. Now-for-then desires, as we saw above, do impact welfare according to SDS, so long as they are coupled with now-for-then beliefs. So *subjective* posthumous satisfaction (if we may call it that), *is* valuable. It is good for you to think your corpse will be treated the way you want it to be treated. The sense in which posthumous satisfactions don't count on SDS is that, after you are dead, there is nothing we can do to affect the welfare-value of your life. So given SDS, we don't benefit the dead by carrying out their wishes about their manner of burial. Since many people think we cannot be harmed or benefited after we are dead, SDS therefore avoids another popular time-related objection to the desire approach. It should be noted that even though SDS deems posthumous satisfactions worthless, the theory does not imply that we ought not to carry out the wishes of the dead. It implies only that if we do have an obligation to carry out the wishes, the obligation derives from some factor other than the welfare-effect on the dead.

SDS avoids The Objection from Remote Desires because only non-remote satisfactions – i.e., cases in which the subject believes the object of his desire is true – count. When the stranger on Parfit's train eventually recovers from illness, Parfit is not made any better off, but not because the desire that the stranger recover isn't "about

Parfit's own life"; rather it is because Parfit has no awareness of the stranger's recovery.<sup>13</sup>

SDS avoids The Problem of Desiring Not To Be Well-Off because it does not in fact imply that things are going well for the man desiring and getting the awful job. True, he does have at least one desire satisfied – that he be badly off. But he has many more and more intense subjective desire frustrations. In order for a job really to be arduous, boring, and unfulfilling, it has to involve much subjective desire frustration.

Subjective Desire Satisfactionism thus in my view avoids or solves the most important problems facing the desire satisfactionist.

Subjective Desire Satisfactionism is a "mental state theory": according to it, how well-off a person is depends solely upon her mental states. Thus, one aspect of traditional desire satisfaction theories that attracted some – that it is a "state of the world" theory rather than a mental state theory – has been abandoned. Though I am convinced that mental state theories are more defensible, I need not take a stand for the purposes of this paper. I'll explain why in the next section, on hedonism.

Subjective Desire Satisfactionism is an unusual form of desire satisfactionism in another respect. According to it, a state of affairs can be good for a subject even though no desire of the subject's is satisfied in it. The subject need only *believe* that the object of his desire obtains. But notice that typical versions of ideal desire satisfactionism – which count only the satisfaction of some class of your hypothetical desires – share this

---

<sup>13</sup> Actually, Parfit's particular case might be solved by the *concurrence* aspect of SDS: maybe the reason the subject is not made better off is that he no longer has the desire. But if we fill in the details of the case so that the subject still has the desire, the belief aspect of SDS will do the trick.

feature as well. On ideal desire satisfactionism, a state of affairs can be good for a person even though no actual desire of the subject's has been satisfied in it.

### 6.3 Hedonism

Simple versions of hedonism are implausible. Those that take pleasure to be a single, uniform sensation fail for two reasons: (1) pleasure is not a single, uniform sensation; and (2) even granting that it is, such a theory entails that people who want, say, peace and quiet instead of these sensations of pleasure, and get it, aren't very well off, despite the fact that they are perfectly satisfied with their lives and are at every moment getting exactly what they want.

A more plausible view about the nature of pleasure, according to which pleasure is, most fundamentally, a propositional attitude rather than a sensation, provides the basis for a more plausible hedonism. Pleasure, on this view, is had when a person "takes pleasure in" some state of affairs, or enjoys it, or is pleased that it is the case. This is the *Attitudinal Theory of Pleasure*. A version of hedonism that makes use of this theory of pleasure implies that the life of the desirer of peace and quite over sensory pleasure very well may be filled with pleasure, though none of it sensory. It will be filled with pleasure so long as the person takes pleasure in the fact that she is getting the peace and quiet.<sup>14</sup>

The Attitudinal Theory of Pleasure should really be called the Attitudinal Theory of Pleasure and Pain, for it treats pain in the analogous way, as a propositional attitude. We might express this attitude when we say that a person is "pained by the

---

<sup>14</sup> This dialectical move is recently made by the hedonist Fred Feldman (2002).

fact that” or “doesn’t like that” such-and-such is the case. It is important to emphasize that on the Attitudinal Theory, there can still be such things as sensations of pleasure and pain. A sensation gets to be a sensation of pleasure (or pain), according to the theory, when the person experiencing the sensation takes pleasure (or pain) in the fact that he is feeling it.<sup>15</sup> On this view, any sensation can be a sensation of pleasure, whether it is the pressure sensation of a massage, the taste sensation of a fine wine, or even a sensation most people would find painful, such as a burning sensation caused by fire. It is a virtue of the Attitudinal Theory of Pleasure that it handles nicely the phenomenon of masochism: a person is a masochist (or at least one sort of masochist) if he takes pleasure in sensations that many of us would be pained by.

A hedonism that makes use of the Attitudinal Theory of Pleasure – “attitudinal hedonism”<sup>16</sup> – requires at least one additional complication: a restriction to *intrinsic* enjoyment (and “disenjoyment”). This is pleasure (or pain) taken in some state of affairs for its own sake, not for what it might lead to. A parallel restriction, recall, was made to the desire theory. Call the resulting theory *Intrinsic Attitudinal Hedonism*, or ‘IAH’ for short.<sup>17</sup> If we assume, as is very natural to do, that some instances of attitudinal pleasure are stronger, or more intense, than others and that these intensities are in principle quantifiable, we can formulate IAH as follows:

---

<sup>15</sup> This thesis must be interpreted as about *intrinsic* attitudinal pleasure (and pain). More on this below.

<sup>16</sup> This is the name Feldman (2002) gives it.

<sup>17</sup> Again, following Feldman (2002).



**Intrinsic Attitudinal Hedonism:**

- (i) Every instance of intrinsic attitudinal pleasure is intrinsically good for its subject.
- (ii) Every instance of intrinsic attitudinal pain is intrinsically bad for its subject.
- (iii) The intrinsic value for its subject of an instance of intrinsic attitudinal pleasure = the intensity of the pleasure.
- (iv) The intrinsic value for its subject of an instance of intrinsic attitudinal pain =  $-(\text{the intensity of the pain})$ .
- (v) The intrinsic value of a life for the one who lives it = the sum of the values of all the instances of intrinsic attitudinal pleasure and pain contained therein.

In light of his rejection of the importance of sensory pleasure, it is not unreasonable to attribute something like attitudinal hedonism to Epicurus.<sup>18</sup> And some version of attitudinal hedonism is held by, or might be held by, or might be part of the theory held by, Parfit (1984, pp. 501-502), later Brandt (1991)<sup>19</sup>, Adams (1999, ch. 3), and Feldman (2002).

Though IAH does not fall prey to objections (1) and (2) above, it does contain a controversial feature: it is a mental state theory.

Some philosophers do not like mental state theories, often for these two reasons:

- (i) a person can be radically deceived about his situation and still lead a good life according to such theories<sup>20</sup>; and (ii) a life filled with only “base pleasures” is still a

---

<sup>18</sup> See Feldman (2004, p. 57).

<sup>19</sup> Brandt calls his theory the “gratification enhancement theory.” It is not unreasonable to suppose that gratification is attitudinal pleasure.

<sup>20</sup> Robert Nozick’s (1974, pp. 42-45) experience machine and Kagan’s (1998, pp. 34-37) deceived businessman each exemplify this objection.

good one (at least according the mental state theory currently under consideration).<sup>21</sup>

My own view is that such objections are misguided. The deceived life and the base life still rank high in terms of welfare, but we are inclined to judge them unfavorably because they rank poorly on other scales on which a life can be measured, such as the scales that measure dignity, or virtue, or achievement. Subjective Desire Satisfactionism and Intrinsic Attitudinal Hedonism, as theories of *welfare*, are therefore not damaged by these “anti-mental-statist” arguments.

But if you disagree, and insist that, say, a perpetual indulger has a lower level of *welfare* than an equally enjoyable but more dignified life, there is a way to revise hedonism (and hence, so say I, desire satisfactionism) to avoid the objection. This line has been suggested by Parfit (1984, pp. 501-502) and Adams (1999, ch. 3) and explicitly taken by Feldman (2002). Parfit (1984, p. 493) distinguishes “objective list” theories of welfare from hedonist and desire-satisfaction theories. According to objective list theories, there are some things – knowledge, engagement in rational activity, mutual love, and awareness of beauty might be contenders (Parfit 1984, p. 502) – that are just good for you to get no matter what your attitude towards them is. Standard hedonist and desire-satisfaction theories are at the other end of the spectrum; according to them, it doesn’t matter what you get, so long as you have the right attitude (pleasure or desire) towards it. Objective list theories may seem unsatisfactory because they make it possible for a person who hates his life through and through nevertheless to have a good one. Desire-satisfaction and hedonist theories seem unsatisfactory to

---

<sup>21</sup> Moore’s (1903, §56) example of the perpetual indulgence in bestiality of course exemplifies this objection.

some because they make it possible for a perpetual indulger in the base to have a great life. Parfit suggests a synthesis (1984, p. 502):

On this view, each side in this disagreement saw only half the truth. Each side put forward as sufficient something that was only necessary. Pleasure with many other kinds of object has no value. And, if they are entirely devoid of pleasure, there is no value in knowledge, rational activity, love, or awareness of beauty. What is of value, or is good for someone, is to have both; to be engaged in these activities, and to be strongly wanting to be so engaged.

Parfit's remarks suggest the following kind of revision to both attitudinal hedonism and desire satisfactionism. Whereas according to the simpler versions of the theories, only the intensity of the pleasure or the desire plays a role in determining welfare value, according to a hybrid hedonist/objective list theory or desire satisfactionist/objective list theory, whether or not the object of the pleasure or the desire is on the objective list also plays a role. One natural way for the theory to go would be to assign a number to every state of affairs, one representing how *worthy of being desired* the state of affairs is, or how *appropriate it is to have pleasure taken in it*.<sup>22</sup> It could be held that the objects of base desires or pleasures have a negative level of pleasure- or desire-worthiness. The theory would then have the implication that satisfying a base pleasure or a base desire actually makes a life go worse. Thus, if one insists that a life of perpetual indulgence is indeed ineligible, one need not abandon the hedonist or desire approach wholesale.

Incidentally, I suspect that those who go in for ideal desire-satisfaction theories are really closet hybrid desire satisfactionist/objective list theorists. A forthright hybrid

---

<sup>22</sup> This is more or less how Feldman (2002) does it. Adams (1999) requires the objects of enjoyment to have *excellence*. Parfit does not give criteria for inclusion on the objective list. The well-known views of J.S. Mill (about the enhanced value of "higher" pleasures) and of G.E. Moore (about the appreciation of the beautiful) also seem to belong in the same family.

theorist of this sort would just come right out and say that it is possible to intrinsically desire the “wrong” things. But ideal desire satisfactionists shy away from such talk, presumably because their theory is not supposed to be so paternalistic.

The cases of radical deception are amenable to similar treatment. This time, it’s not pleasure-worthiness or desire-worthiness that counts, but *truth*. On the revised theory, pleasure taken in, or the subjective satisfaction of desires for, *true* states of affairs enhances the value of the pleasure or the subjective satisfaction. This way, a deluded life lived in the Matrix isn’t as good as the corresponding real life, even though the two lives are indiscernible from the inside. Mental-statism is abandoned.

I am describing ways in which both hedonism and the desire theory could either incorporate or reject a truth requirement. Hedonism as it is normally understood can be strengthened into a state of the world theory, and desire satisfactionism as it is normally understood can be weakened into a mental state theory (like SDS). My aim in the paper is to establish a link between hedonism and desire satisfactionism, but the mental-statism debate might be precisely the debate that divides the two camps. If it is, then perhaps what I am saying to each camp is something like the following. You desire theorists wedded to state-of-the-worldism: your best theory is in fact equivalent to a form of hedonism that you should like just as much. And you hedonists wedded to mental-statism: your best theory is in fact equivalent to a form of desire satisfactionism that you should like just as much.

We could go on modifying. Parfit (1984, p. 497) argues against summative theories (such as SDS and IAH) and in favor of “global” theories. Though I am not persuaded by Parfit’s argument (the drug-addiction argument), we need not settle the

matter here. This is because the change Parfit suggests – viz., that the desire theory count only “global desires,” or desires about one’s life as whole – can likewise be made to the hedonist theory. Since, again, my principal aim in this paper is to demonstrate the link between hedonism and desire satisfactionism, and since parallel changes can be made to each type of theory, I will set aside both the mental-statism debate and the summative/global debate. Whatever the truth is in the matter, revise IAH and SDS accordingly. You’ll still be left with numerically identical theories.

In what follows, I will assume for simplicity’s sake that we have settled on IAH and SDS.

#### **6.4 The Motivational Theory of Pleasure**

We have now, before us, what seem to be two rival theories of the good life: a form of hedonism (IAH) and a form of desire satisfactionism (SDS). How do we choose between them? Luckily, we don’t have to. For we have only one theory on our hands: IAH *is* SDS . This claim rests upon the theory of pleasure that I favor, the *Motivational Theory of Pleasure*, which reduces pleasure to desire.

You might think that the reduction of intrinsic attitudinal pleasure to intrinsic desire should go something like this: S is intrinsically pleased at t that p is true iff S intrinsically desires at t that p be true, and p is true. But this won’t do. As a theory about the conditions under which a person is pleased that something is the case, whether or not the thing really is the case is irrelevant. I might strongly intrinsically desire that the stranger I met on the train get well, and he might in fact get well, but if I’m not aware of it, I certainly won’t be pleased about it. So the truth requirement in a



motivational theory of attitudinal pleasure (in combination with the desire requirement) is not sufficient. The truth requirement is not necessary either. We can be pleased about false states of affairs, as when Al Gore, for a time, took delight in winning the election, even though he didn't actually win.

Actually, this idea of false pleasures faces an annoying linguistic difficulty: 'pleased that' is, or appears to be, "factive": statements of the form 'S is pleased that p' appear to entail p. In other words, truth appears to be necessary for attitudinal pleasure after all. But the "factivity" of 'pleased that' is not obvious. Suppose Smith believes he won the election and is ecstatic as a result. Suppose also that in fact he didn't win and the rest of us know this. I concede that in this context there is something wrong with me saying to you, as we observe Smith, "Wow, Smith is really pleased that he won." But I think it is not at all clear whether the defect in this sentence is semantic (in that the sentence is just false, no matter how pleased Smith is) or pragmatic (in that the sentence, while literally true, is, for some reason or other, pragmatically unassertable).

Either way, we're ok. If the defect is merely pragmatic, then the claim that truth is not necessary for attitudinal pleasure stands. On the other hand, if attitudinal pleasure really does entail truth, then we must acknowledge that the attitude expressed in sentences of the form 'S is pleased that p' is not a "purely psychological" attitude (just as, e.g., knowledge is not a purely psychological attitude). But even if this is true, clearly there is a purely psychological attitude in the vicinity. Suppose Jones believes he won the election and is ecstatic as a result. Suppose also that Jones did win and we all know it. In this case we *can* say, "Jones is pleased that he won the election." Now, clearly, there is some one psychological attitude that is shared by both Smith and Jones

(or that is shared by both Smith and his counterpart in worlds doxastically indiscernible from Smith's perspective but in which Smith wins). Call this attitude attitudinal pleasure\*. Since Intrinsic Attitudinal Hedonism means to recognize the possibility and even the value of false pleasures, it is actually attitudinal pleasure\* that the attitudinal hedonist has in mind. This purely psychological attitude is what the Motivational Theory of Pleasure is supposed to be about as well. I will henceforth drop the asterisk and simply use 'is pleased that' and its variants to mean the purely psychological kind of attitudinal pleasure.

Although truth is not necessary for attitudinal pleasure, it seems pretty clear that *belief* is. That is, in order for someone to take pleasure in some proposition, she must believe that the proposition is true. This suggests the following improvement upon the previous version and is indeed the theory I endorse, which I will call 'MTP' for short:

**MTP:** S is intrinsically pleased at t that p iff S intrinsically desires at t that p and S believes at t that p.<sup>23</sup>

MTP takes desire and belief as basic psychological attitudes and reduces pleasure to them. But why think, between desire and pleasure, that desire is the basic attitude?

Why not take pleasure as basic and hold that desire is reducible to it?<sup>24</sup>

This is because it can be shown that attitudinal pleasure is not a basic attitude. We know it's not basic because, as noted above, having this attitude entails having another one: belief. But there can be no such necessary connections between distinct

---

<sup>23</sup> A more complete formulation of the theory would include variables representing *intensity* of pleasure and the way intensity of pleasure depends upon intensity of desire (or of desire and belief).

<sup>24</sup> An impressive catalog of alleged adherents of views along this line can be found in Fehige (2001); Fehige himself also defends a reduction of desire to pleasure.

attitudes. Given MTP, the necessary connection between pleasure and belief is not one between distinct attitudes, so the problem is avoided; the connection is no more mysterious than that between bachelorhood and maleness.

Now what about MTP itself? Is it plausible? Doesn't it sometimes happen that we get what we want, are aware that we have gotten what we want, but fail to enjoy it? For example, I recently wanted Froot Loops, a sugary cereal I loved as a kid. Then I got some Froot Loops. But I hated them – way too sweet! Also, doesn't it sometimes happen that I enjoy things I never desired? For example, as I was working on this paper, my wife approached me and unexpectedly began massaging my shoulders. This is something I wasn't thinking at all about when it occurred, but I took great pleasure in it.

I think neither argument succeeds. In the Froot Loops case, I began with a desire for a certain taste, but once I got it, I lost my desire. Indeed, I didn't take a second bite and even wanted to spit out the first. What explains this other than the fact that I no longer wanted to be tasting the Froot Loops. Since the belief and desire never did overlap, MTP does not imply that I enjoyed the Froot Loops. Yes, it can be that a person gets what she wants without enjoying it, so long as she loses the desire the instant she gets it.

In the surprise massage case, as soon as I became aware of the wonderful sensations, I instantly desired to be feeling them. Suppose my wife had asked, "Shall I keep it up?" I would have answered unhesitatingly in the affirmative. This, of course, is because I *wanted* to be feeling the sensations I was then feeling (even though I had no thoughts about such sensations one way or the other before she began). MTP therefore

does imply, correctly, that I enjoyed the massage. Yes, it can be that a person enjoys something he didn't have a desire for, so long he forms a desire for it the instant he starts getting it.

The two arguments just considered fail to appreciate the temporal index in MTP. Other arguments fail to appreciate the intrinsicity requirement. Last night I flossed my teeth. I did it because I wanted to do it. As I continued to floss I continued to want to be flossing; that explains my continued flossing. The whole time I flossed I also believed I was flossing. But I hereby testify to this: at no time during the process did I take any pleasure in the fact that I was flossing. So is this case of a desire and a belief without a pleasure a counterexample to MTP? No, for my desire to floss was merely extrinsic – I had it only for the sake of the health of my teeth (the desire for which, incidentally, is probably also extrinsic). It is much harder to come up with a case of a person intrinsically wanting and (thinking he is) getting something while failing to enjoy it, and vice versa. If he is enjoying it, he must be liking it. If he is liking it, he must be “into it.” But this is just to say that he is for it, that he’s “pro” it. But this is just what it is to desire it.

## **6.5 Two Theories or One?**

Two theories of welfare are genuinely distinct theories if they disagree about the value of some possible lives. Though the converse arguably does not hold – two distinct theories may be necessarily extensionally equivalent – no two theories are ever intensionally equivalent. If “they” were, “they” would be just one theory; for to be

intensionally equivalent is to express the same proposition, and theories just are propositions.

IAH and SDS are not only necessarily extensionally equivalent, they are the same theory. IAH says that attitudinal pleasure constitutes welfare; SDS says that subjective desire satisfaction constitutes welfare. MTP tells us that attitudinal pleasure just is subjective desire satisfaction. Since MTP is supposed to be analytic, it follows that the above claims of IAH and SDS are in fact just one proposition. Hedonism and desire satisfactionism, once the kinks are worked out, turn out to be the same theory. That we have arrived at this theory via two independent avenues gives us added reason to think the theory is true.



## BIBLIOGRAPHY

- Adams, Robert Merrihew, *Finite and Infinite Goods* (Oxford: Oxford University Press, 1999).
- Alston, William, "Pleasure," in Paul Edwards (ed.) *The Encyclopedia of Philosophy* (New York: Macmillan Publishing Co. & The Free Press, 1967): 341-347.
- Aquinas, Thomas, *Summa Theologica*.
- Arneson, Richard J., "Human Flourishing Versus Desire Satisfaction," *Social Philosophy and Policy* 16 (1999): 113-142.
- Aristotle, *Nicomachean Ethics*, Roger Crisp (trans. and ed.) (Cambridge: Cambridge UP, 2000).
- Bentham, Jeremy, *An Introduction to the Principles of Morals and Legislation*, J.H. Burns and H.L.A. Hart (eds.) (Oxford: Oxford UP, 1996).
- Brandt, Richard B., "Rationality, Egoism, and Morality," *Journal of Philosophy* 69 (1972): 681-97.
- Brandt, Richard B., *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979).
- Brandt, Richard B., "Two Concepts of Utility," in Miller and Williams (eds.) *The Limits of Utilitarianism* (Minneapolis, MN: University of Minnesota Press, 1982), pp. 169-185.
- Brandt, Richard B., "Overvold on Self-Interest and Self-Sacrifice," *Journal of Philosophical Research* 16 (1991): 353-363.
- Bricker, Phillip, "Prudence," *Journal of Philosophy* 77 (1980): 381-401.
- Bykvist, Krister, "Sumner On Desires and Well-Being," *Canadian Journal of Philosophy* (2002): 475-490.
- Carson, Thomas L., "The Desire-Satisfaction Theory of Welfare: Overvold's Critique and Reformulation," in John Heil, ed., *Rationality, Morality, and Self-Interest: Essays Honoring Mark Carl Overvold* (Lanham, MD: Rowman & Littlefield, 1993): pp. 233-246.
- Carson, Thomas L., *Value and the Good Life* (Notre Dame, IN: University of Notre Dame Press, 2000).

- Davis, Wayne, "A Theory of Happiness," *American Philosophical Quarterly* 18 (1981): 111-120.
- Davis, Wayne, "A Causal Theory of Enjoyment," *Mind* 91 (1982): 240-256.
- Edwards, Rem B., *Pleasure and Pains: A Theory of Qualitative Hedonism* (Ithaca: Cornell University Press, 1979).
- Fehige, Christoph, "Instrumentalism," in E. Millgram (ed.), *Varieties of Practical Reasoning* (Cambridge, MA: MIT Press, 2001).
- Feldman, Fred, "Some Puzzles About the Evil of Death," *The Philosophical Review* 100 (1991): 205-227.
- Feldman, Fred, "Two Questions about Pleasure," in Fred Feldman, *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy* (Cambridge: Cambridge UP, 1997): 82-105.
- Feldman, Fred. "Basic Intrinsic Value," *Philosophical Studies* 99 (2000): 319-346.
- Feldman, Fred, "Hedonism," in Lawrence C. Becker and Charlotte B. Becker (eds.) *The Encyclopedia of Ethics* (New York: Routledge, 2001): 662-669.
- Feldman, Fred, "The Good Life: A Defense of Attitudinal Hedonism," *Philosophy and Phenomenological Research* 65 (2002): 604-627.
- Feldman, Fred, *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism* (Oxford: Clarendon Press, 2004).
- Gibbard, Allan, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge: Harvard University Press, 1990).
- Glover, Jonathan, *Causing Death and Saving Lives* (Harmondsworth, NY: Penguin Books, 1977)
- Gosling, Justin, "Pleasure," in Lawrence C. Becker and Charlotte B. Becker (eds.) *The Encyclopedia of Ethics* (New York: Routledge, 2001): 1326-1329.
- Griffin, James, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Clarendon Press, 1986).
- Hall, Richard J., "Are Pains Necessarily Unpleasant?," *Philosophy and Phenomenological Research* 49 (1989): 643-659.
- Hare, R.M., *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981).

- Hooker, Brad, "Mark Overvold's Contribution to Philosophy," *Journal of Philosophical Research* 16 (1991): 333-344.
- Hooker, Brad, "Subjectivism," in Becker and Becker (eds.) *The Encyclopedia of Ethics* (New York: Routledge, 2001).
- Hurka, Thomas, "Perfectionism," in Becker and Becker (eds.) *The Encyclopedia of Ethics* (New York: Routledge, 2001).
- Kagan, Shelly, "The Limits of Well-Being," *Social Philosophy and Policy* 9 (1992): 169-189, reprinted in Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul (eds.) *The Good Life and the Human Good* (Cambridge: Cambridge UP, 1992), pp. 169 – 189.
- Kagan, Shelly, *Normative Ethics* (Boulder, CO: Westview Press, 1998).
- Kenny, Anthony, *Action, Emotion, and Will* (Routledge & Kegan Paul, 1963).
- Korsgaard, Christine M., *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).
- Kraut, Richard, "Desire and the Human Good," *Proceedings and Addresses of the American Philosophical Association* 68 (1994): 39-54.
- Loeb, Don, "Full Information Theories of the Good," *Social Theory and Practice* 21 (1995): 1-30.
- Matthews, Gareth B. and Cohen, S. Marc, "Wants and Lacks," *The Journal of Philosophy* 64 (1967): 455-456.
- Mill, J.S., *Utilitarianism*, Roger Crisp (ed.) (Oxford: Oxford UP, 1998).
- Moore, G.E., *Principia Ethica*, Thomas Baldwin (ed.) (Cambridge: Cambridge UP, 1993).
- Noggle, Robert, "Integrity, the Self, and Desire-Based Accounts of the Good," *Philosophical Studies* 96 (1999): 303-331.
- Nozick, Robert, *Anarchy, State and Utopia* (New York: Basic Books, 1974).
- Overvold, Mark Carl, "Self-Interest and the Concept of Self-Sacrifice," *Canadian Journal of Philosophy* 10 (1980): 105-118.
- Parfit, Derek, *Reasons and Persons* (Oxford: Clarendon Press, 1984).

- Perry, David, *The Concept of Pleasure* (The Hague: Mouton, 1967).
- Rachels, Stuart, "Is Unpleasantness Intrinsic to Unpleasant Experiences?" *Philosophical Studies* 99 (2000): 187-210.
- Rachels, Stuart, "Six Theses About Pleasure," *Philosophical Perspectives* 18 (2004): 247-267.
- Railton, Peter, "Facts and Values," *Philosophical Topics* 14 (1986): 5-31.
- Rawls, John, *A Theory of Justice* (Cambridge, MA: Belknap, 1971).
- Rosati, Connie S., "Person, Perspectives, and Full Information Accounts of the Good," *Ethics* 105 (1995): 296-325.
- Ryle, Gilbert, "Pleasure," *Proceedings of the Aristotelian Society*, supplementary volume 28 (1954): 135-146.
- Scanlon, T.M., *What We Own to Each Other* (Cambridge, MA: Harvard University Press, 1998).
- Schueler, G.F., *Desire: Its Role in Practical Reason and the Explanation of Action* (Cambridge, MA: The MIT Press, 1995).
- Schwartz, Thomas, "Human Welfare: What It Is Not," in Miller and Williams (eds.) *The Limits of Utilitarianism* (Minneapolis, MN: University of Minnesota Press, 1982), pp. 195-206.
- Sen, Amartya, *Choice, Welfare, and Measurement* (Cambridge, MA: MIT Press, 1982).
- Sidgwick, Henry, *The Methods of Ethics*, seventh edition (Macmillan, 1907).
- Sobel, David, "Full Information Accounts of Well-Being," *Ethics* 104 (1994): 784-810.
- Sobel, David, "Varieties of Hedonism," *Journal of Social Philosophy* 33 (2002): 240-256.
- Spencer, Herbert, *The Principles of Psychology* (New York: D. Appleton and Company, 1871).
- Sprigge, Timothy, *The Rational Foundations of Ethics* (London: Routledge and Kegan Paul, 1988).
- Sumner, L.W., *Welfare, Happiness, & Ethics* (Oxford: Clarendon Press, 1996).

- Sumner, L.W., "Review of Fred Feldman, *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*," *Ethics* 109 (1998): 176-179.
- Tännsjö, Torbjörn, *Hedonistic Utilitarianism* (Edinburgh: Edinburgh University Press, 1998).
- Velleman, J. David, "Brandt's Definition of 'Good'," *The Philosophical Review* 97 (1988): 353-371.
- Velleman, J. David, "Well-being and time," *Pacific Philosophical Quarterly* 72 (1991): 48-77.



